Individual-specific networks

Federico Melograna

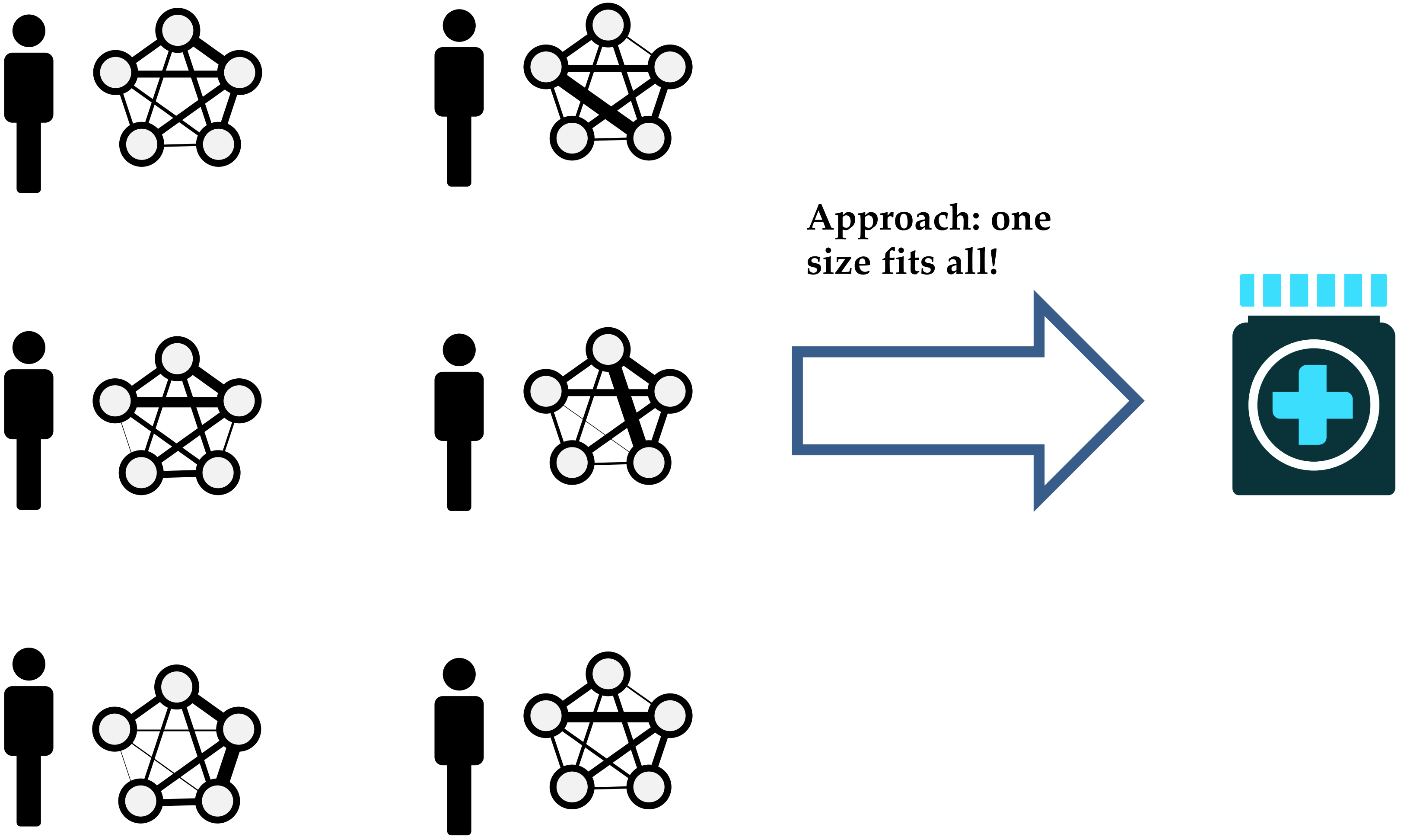
06/12/2022

AGENDA

1. Introduction to Personalized medicine
2. Rationale: networks and networks’ attributes
3. Individual-specific networks
4. Most interesting approaches
5. Comparison
6. Direction and future applications
7. Conclusion

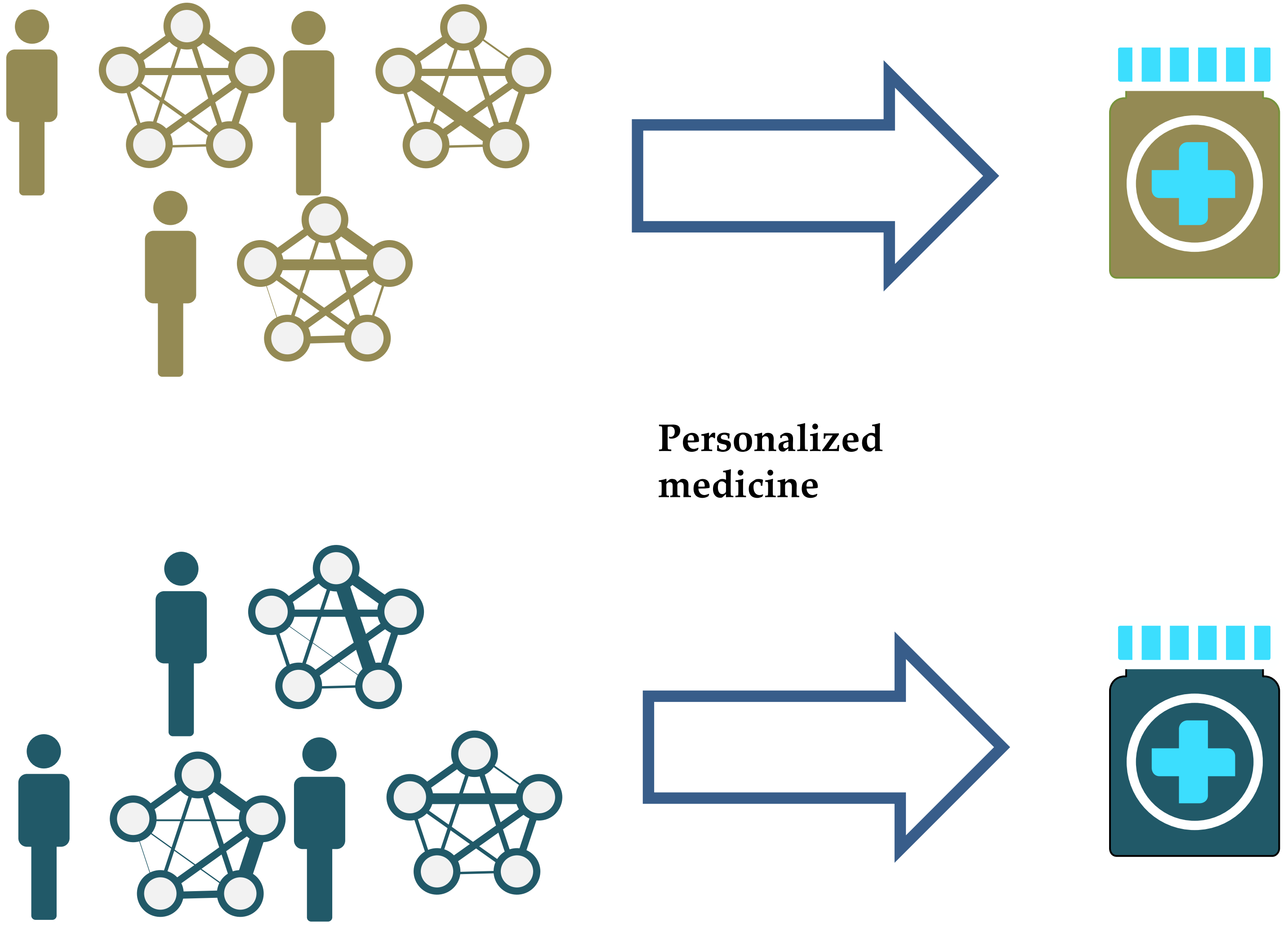
Personalized Medicine

SITUATION



Approach: one size fits all!

PERSONALIZED MEDICINE

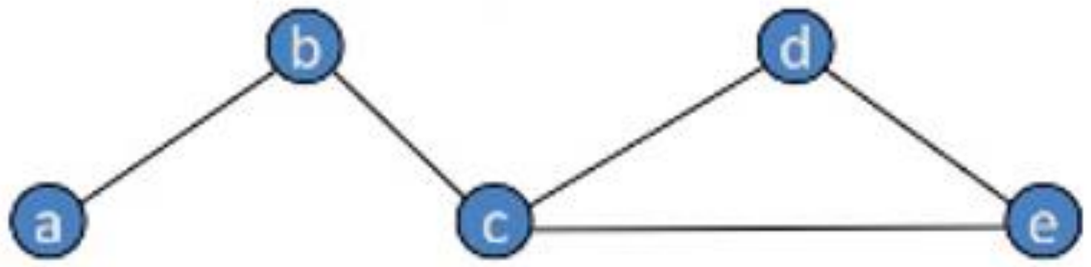


Personalized medicine

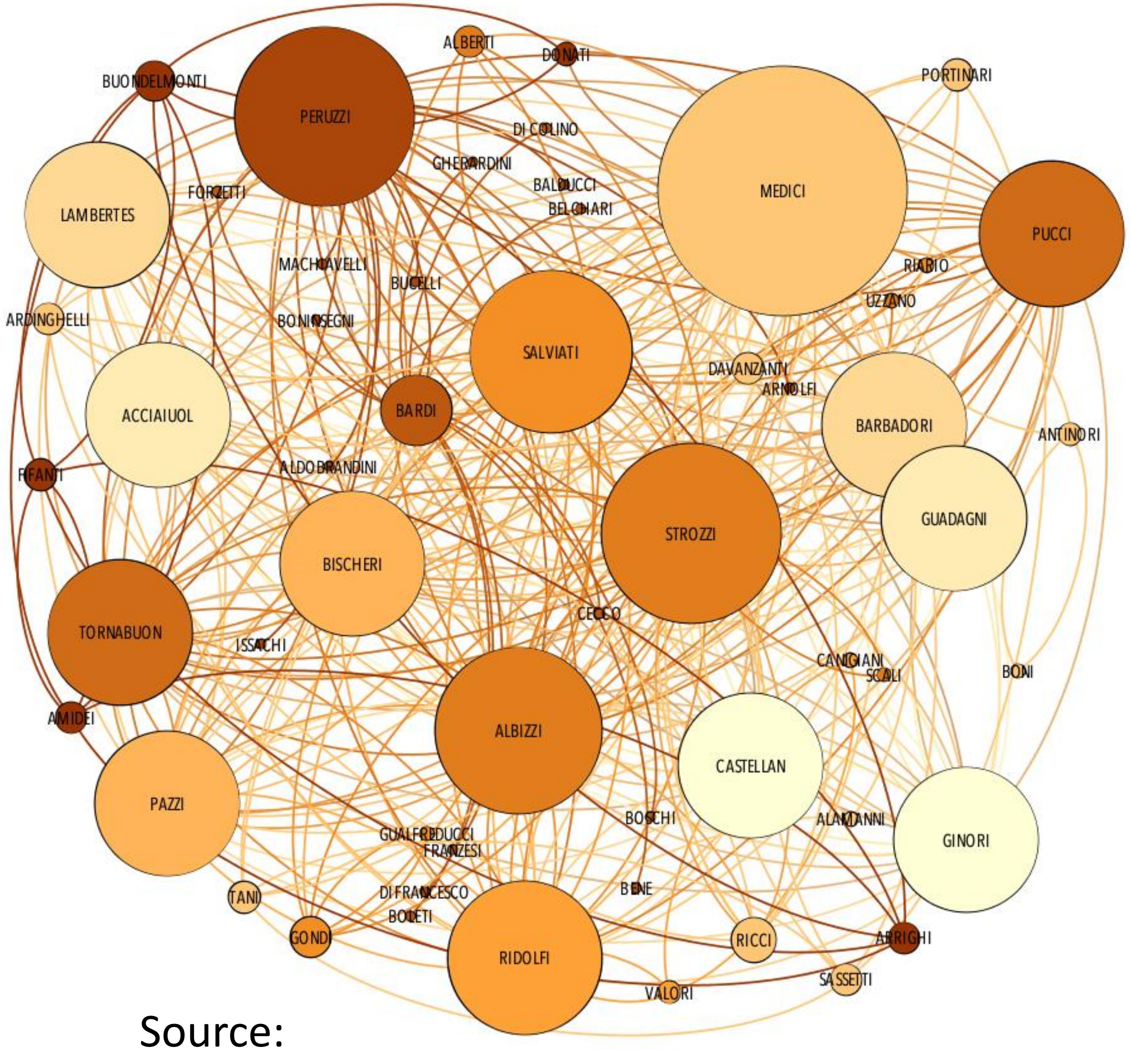
NETWORK DEFINITION

Network theory

- **Vertices** V : Entities of the same type (i.e. genes, proteins, SNPs) = nodes, vertices, points
- **Edges** E: connections between two nodes (association, correlation) = arcs, lines, ties
- **Graph**: a set $G = (V, E)$ of vertices and edges. V is a finite, non-empty set of p nodes and E is a subset of $V \times V$ containing pairs of connected nodes $e_{ij} \coloneqq (v_i, v_j)$
- **Module** (=subnetwork) G' : limited and strongly associated sets of nodes and relative edges, A module $G' = (V', E')$ is a network such that $V' \subseteq V$ and $E' \subseteq E$.



NETWORK VISUALIZATION

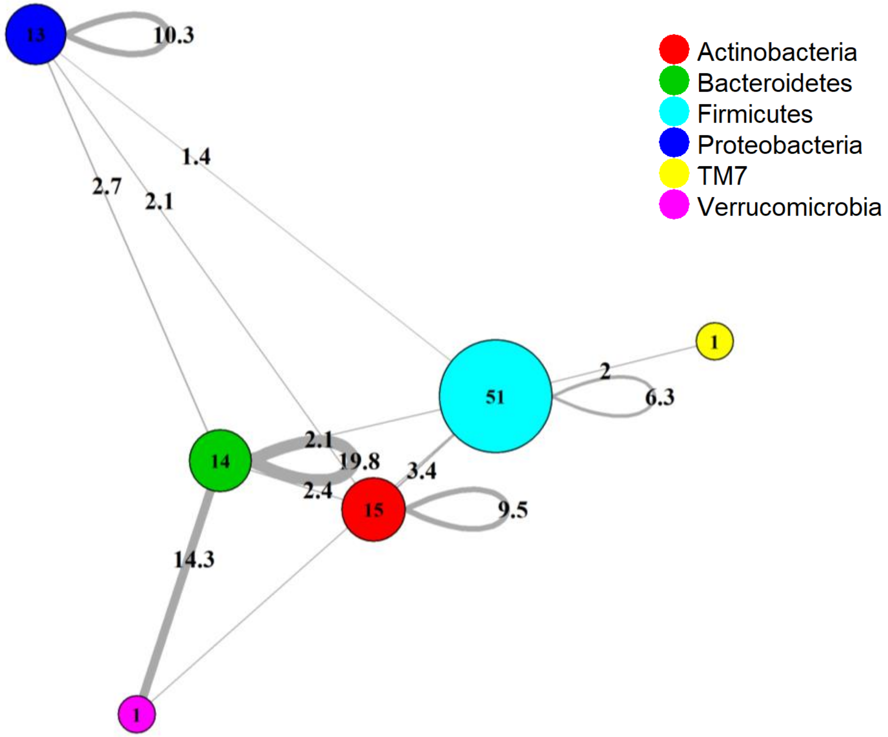


Source: https://studentwork.prattsi.org/infovis/labs/visualizing-florentine-family-networks/

NETWORK DESCRIPTORS

A network can be:

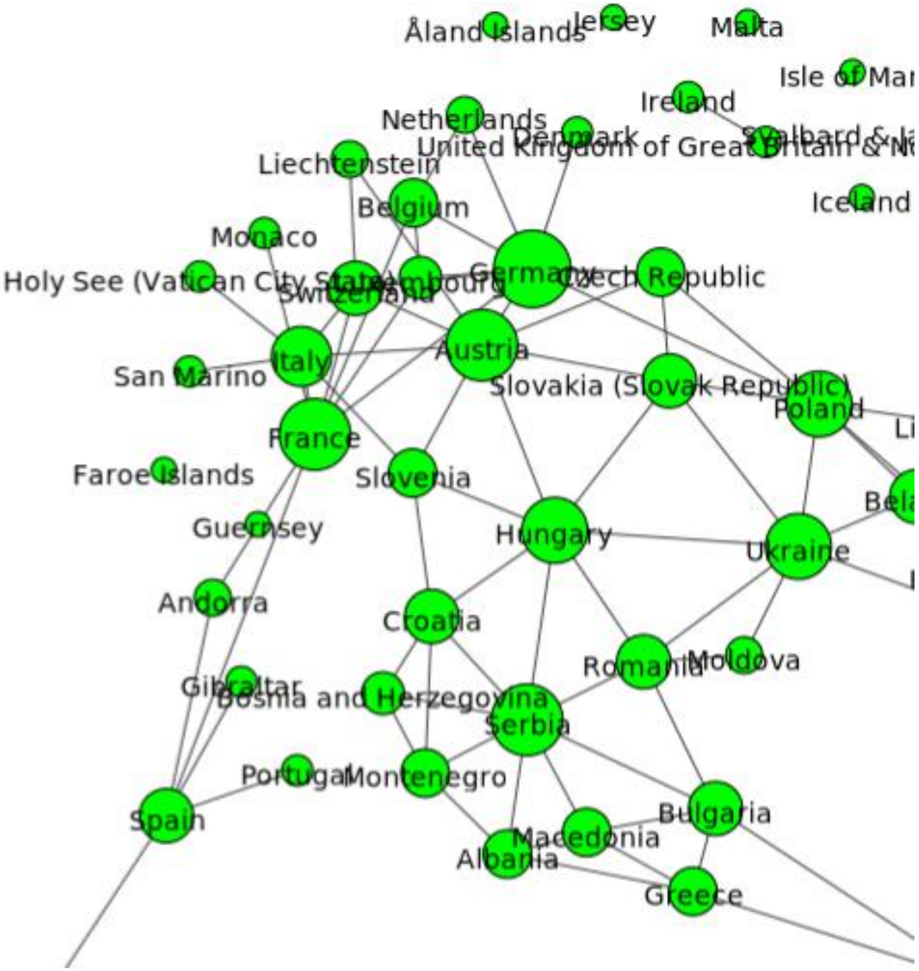
- **Weighted** or Unweighted



NETWORK DESCRIPTORS

A network can be:

- Weighted or **Unweighted**

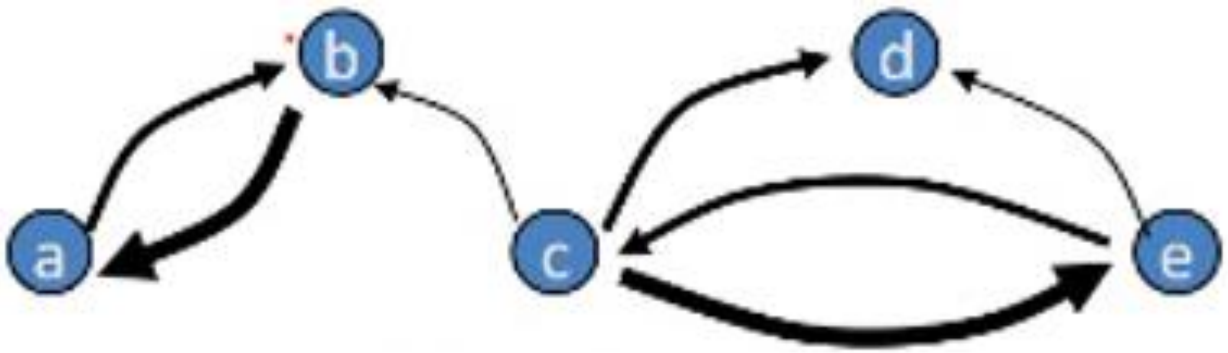


Source: https://i.redd.it/7y5s15gs0cw61.png

NETWORK DESCRIPTORS

Network theory

- **Directed** or Undirected. In directed graphs (digraph), each edge has a direction such that $e_{ij} \neq e_{ji}$

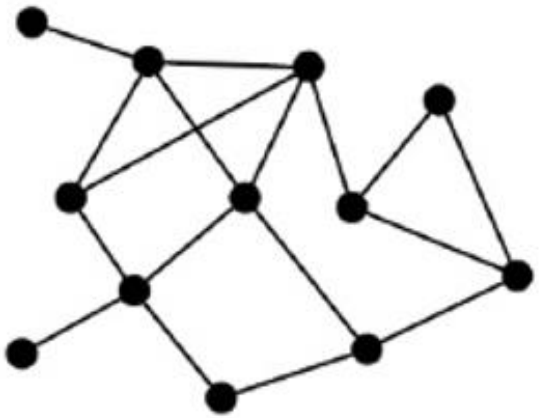


NETWORK PROPERTIES

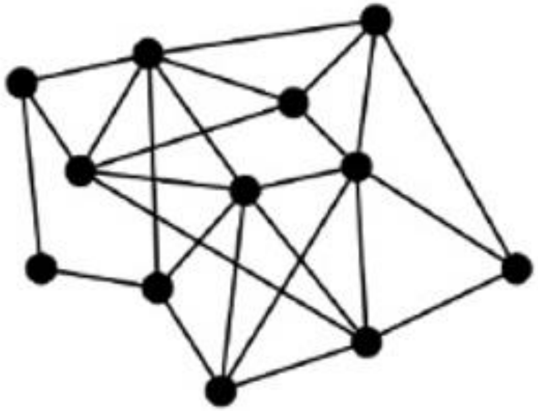
Graphs statistics

- **Density**: number of edges in the number / number of possible edges

- Dense vs sparse



low density: 25%

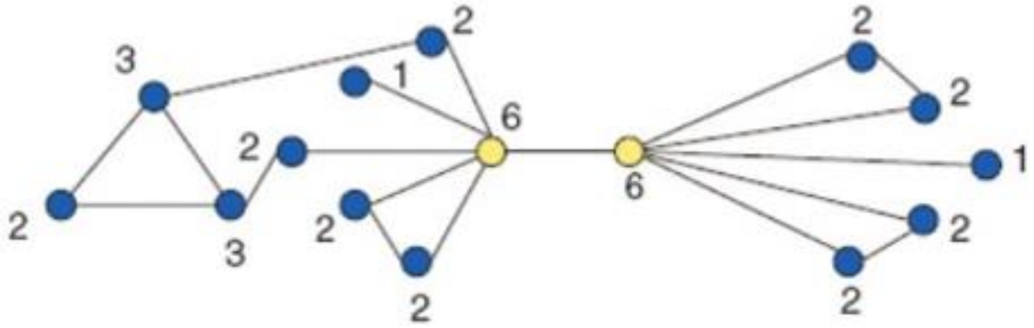


high density: 39%

NETWORK PROPERTIES

Graphs statistics

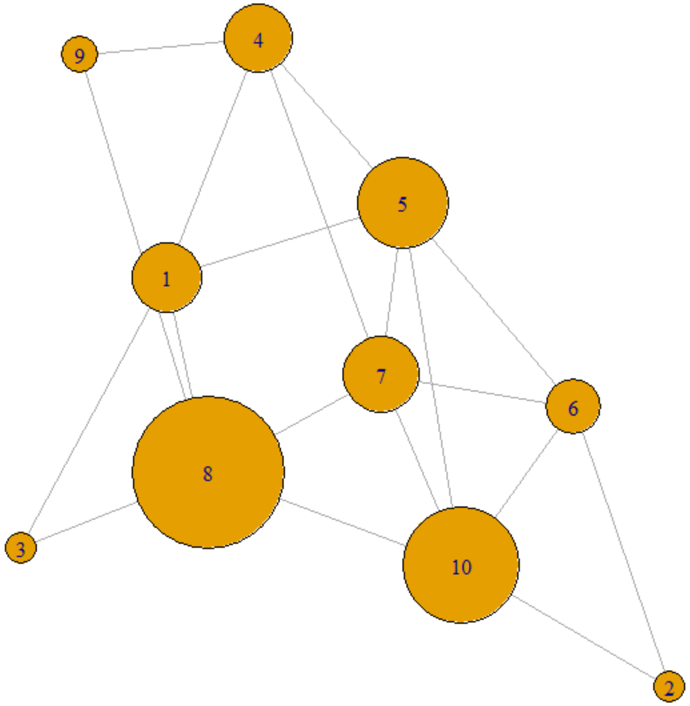
- **Degree**: avg number of edges for each node



NETWORK PROPERTIES

Graphs statistics

- **Betweenness**: number of shortest paths going through a vertex

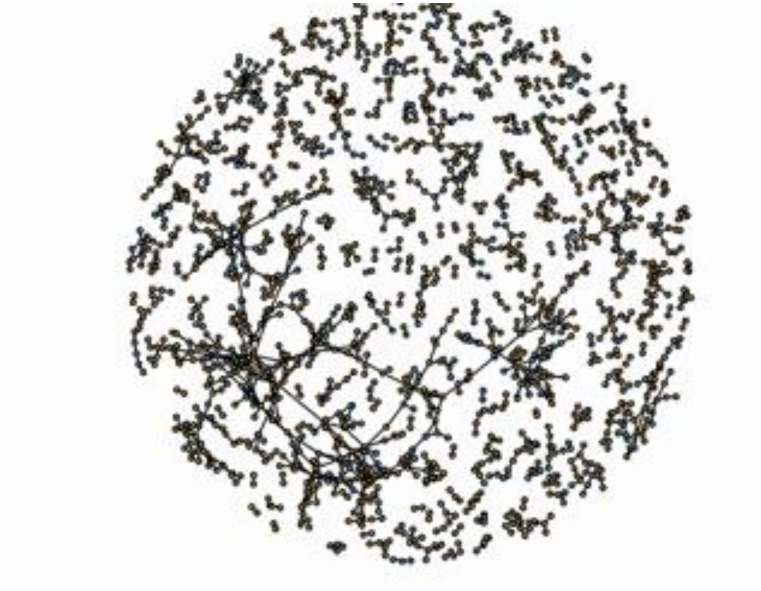


And many more. Ref: https://igraph.org/

INDIVIDUAL-SPECIFIC NETWORK

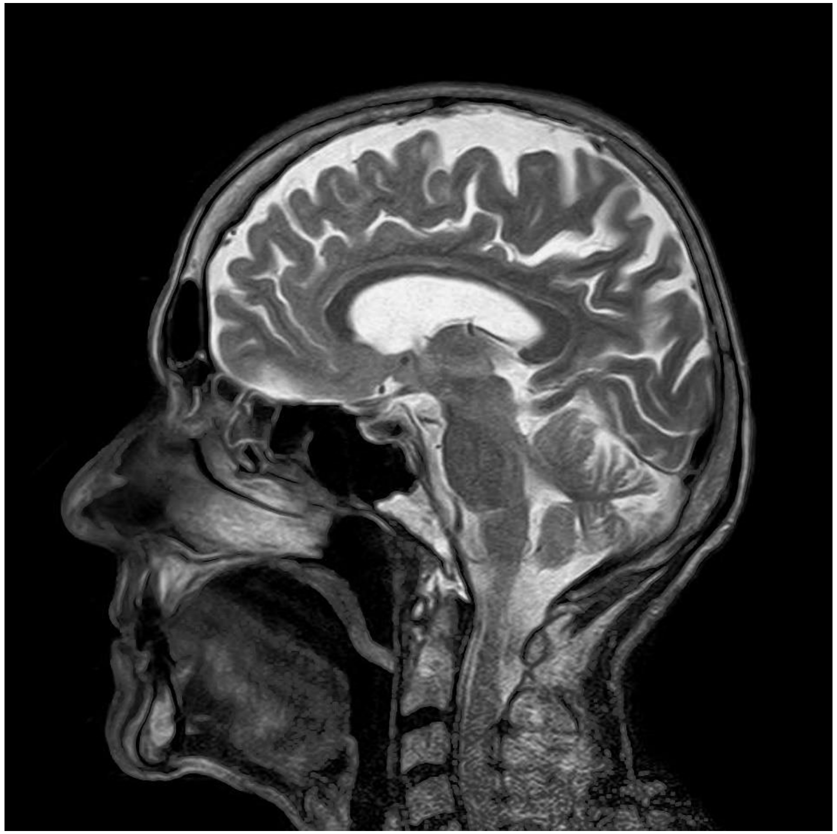
Individual-specific networks (ISNs) are networks in which either ***nodes*** or ***edges*** are individual-specific. ISNs can refer to the following types of networks:

- Weighted or Unweighted
- Directed or Undirected
- Built on multiple measurement or single measurement



INDIVIDUAL-SPECIFIC NETWORK

- Built on **multiple** measurement or single measurement



- We focus on **single** measurement

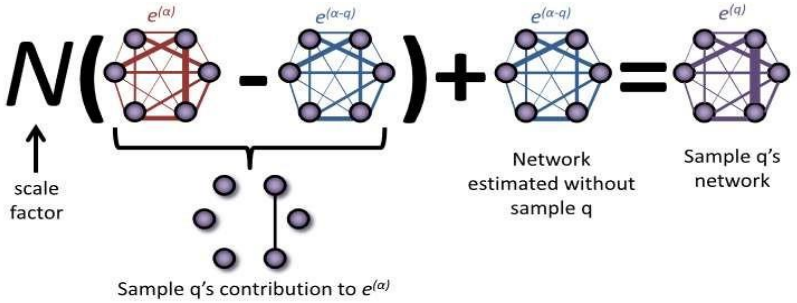
INDIVIDUAL-SPECIFIC NETWORKS (1/2)

1 What?

- Networks that refer to co-occurrence, association, interaction
- In the literature:
  - ➢ Usually based on multiple measurements for the same individual (e.g. neurosciences)
  - ➢ Individual-specific nodes on a fixed edge template common to all individuals (e.g., protein-protein interaction network, inferred gene regulatory network)
- Recently:
  - ➢ **Individual-specific edges** (individual-specific node information available or not)

INDIVIDUAL-SPECIFIC NETWORKS (2/2)

2 How?

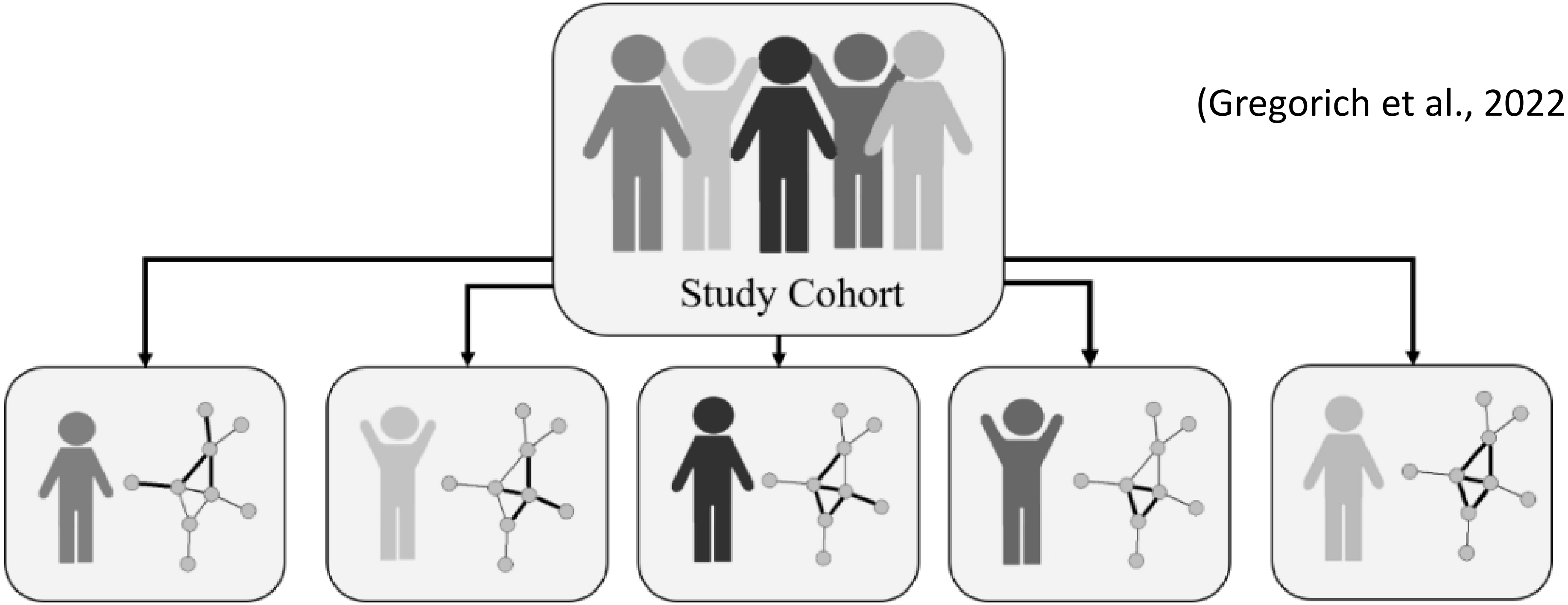


doi: 10.1016/j.isci.2019.03.021
(Kuijjer et al. 2019)

3 Why?

- Networks derived from a collection of individuals can be seen as models for an "average" individual
- Translating network interpretation strategies from pop. to indiv. assumes extrapolations can be made to the level of the individual
- Individual-specific networks allow focusing on each individual and its specific dynamics and associations.

ISN SUMMARY



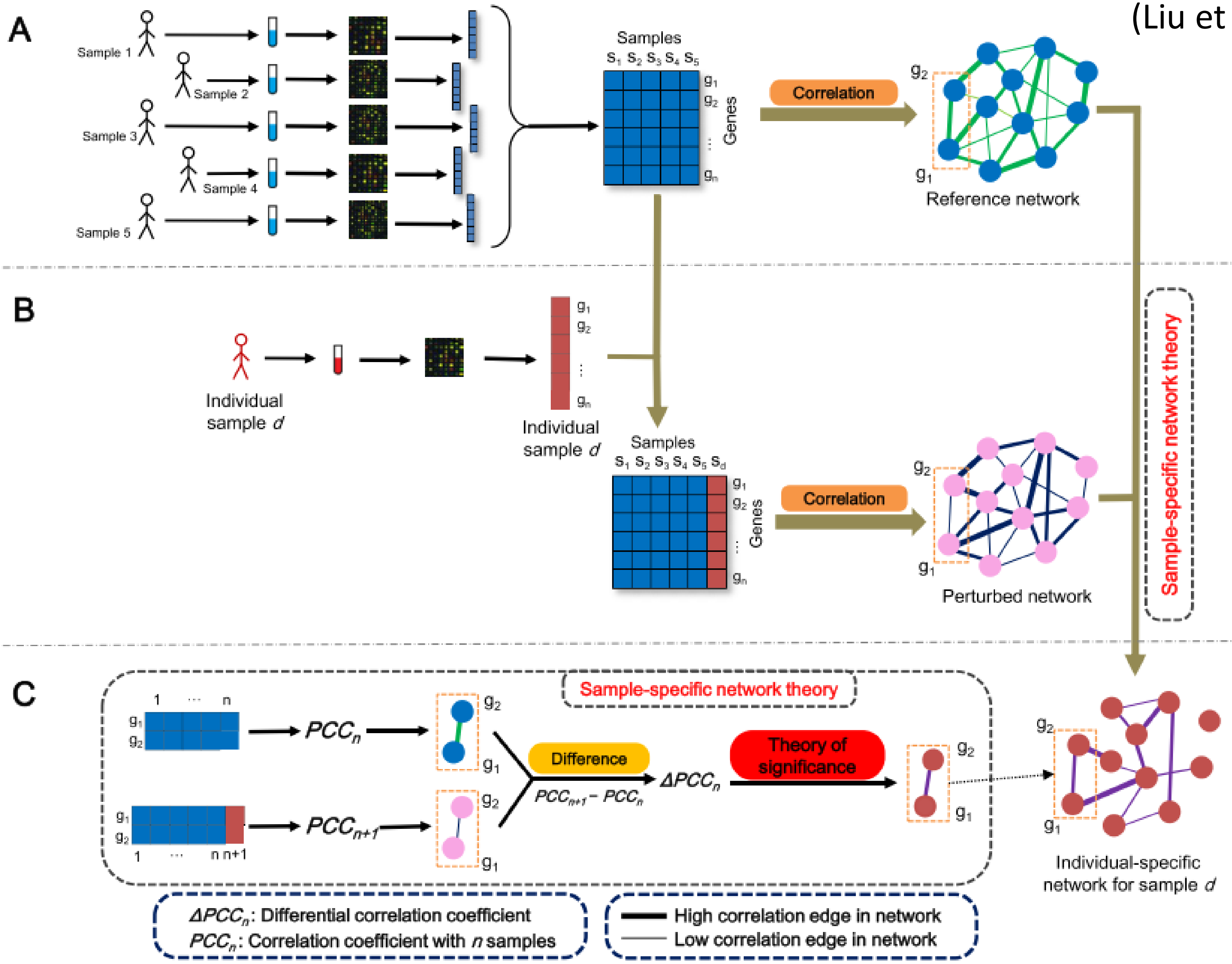
(Gregorich et al., 2022)

For **individual-specific networks**, we assume that for each individual s $(s = 1, \dots, N)$ a unique network $G_s = (V_s, E_s)$ exists, where N is the number of individuals within the study cohort.

In this presentation we cover ISNs built on a **single sample** per-individual where the individual-specificity is on the edges!

Sample-specific networks SSN (1/4): pipeline

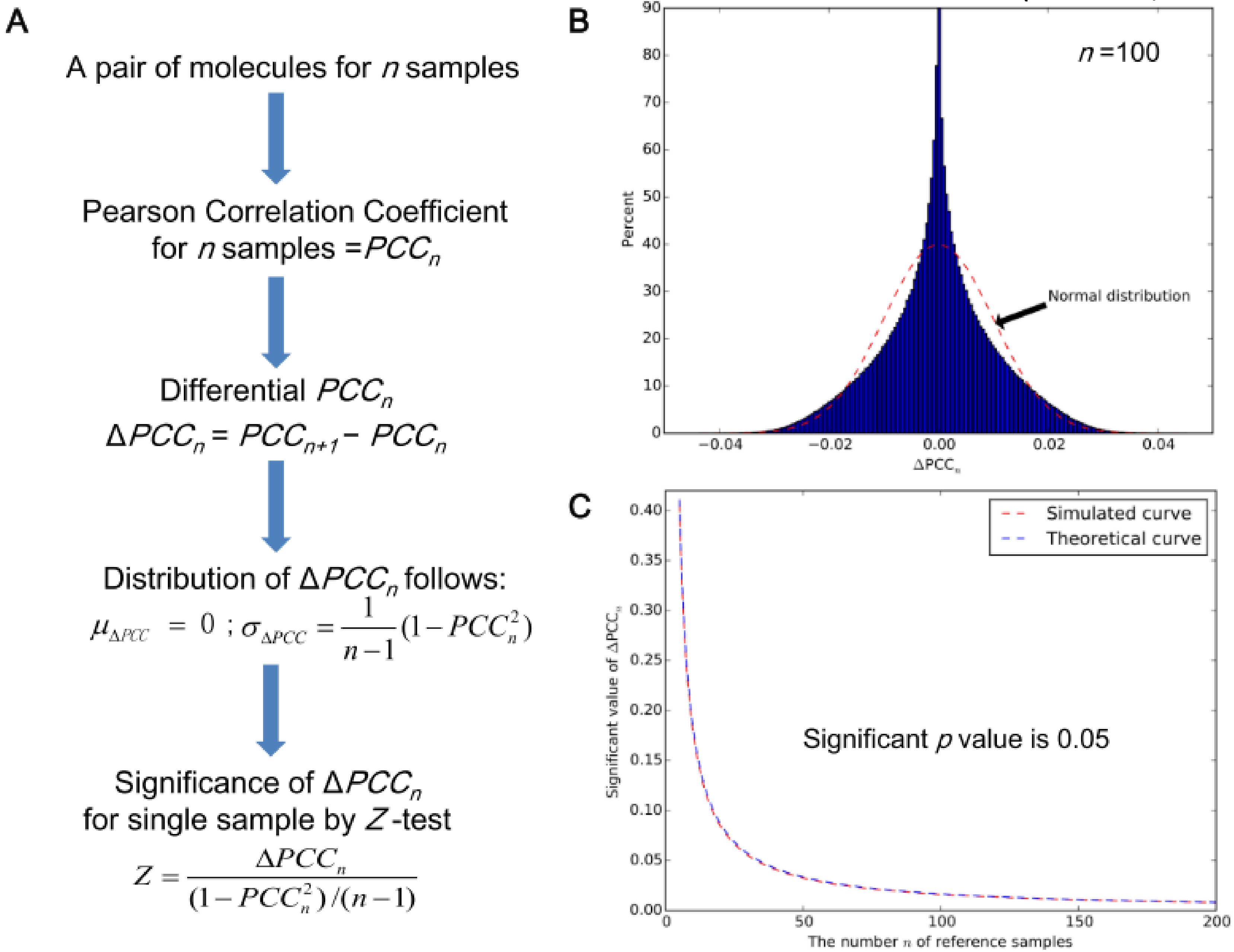
(Liu et al., 2016)



Sample-specific networks SSN (2/4): significance

Individual network

(Liu et al., 2016)



A

A pair of molecules for n samples

Pearson Correlation Coefficient for n samples $= PCC_n$

Differential PCC_n

$\Delta PCC_n = PCC_{n+1} - PCC_n$

Distribution of ΔPCC_n follows:

$$\mu_{\Delta PCC} = 0 \; ; \; \sigma_{\Delta PCC} = \frac{1}{n-1}(1 - PCC_n^2)$$

Significance of ΔPCC_n for single sample by Z-test

$$Z = \frac{\Delta PCC_n}{(1 - PCC_n^2)/(n-1)}$$

Sample-specific networks SSN (3/4): comments

(Liu et al., 2016)

1 How

- Based on a reference network: control samples
- Gives a measure of significance (p-value for the edges)
- Test statistic: $Z = \frac{\Delta PCC_n}{(1-PCC_n^2)/(n-1)}$
- Build a **perturbed** network ΔPCC_n = difference in correlation adding a case sample: only calculated perturbation

2 Caveats

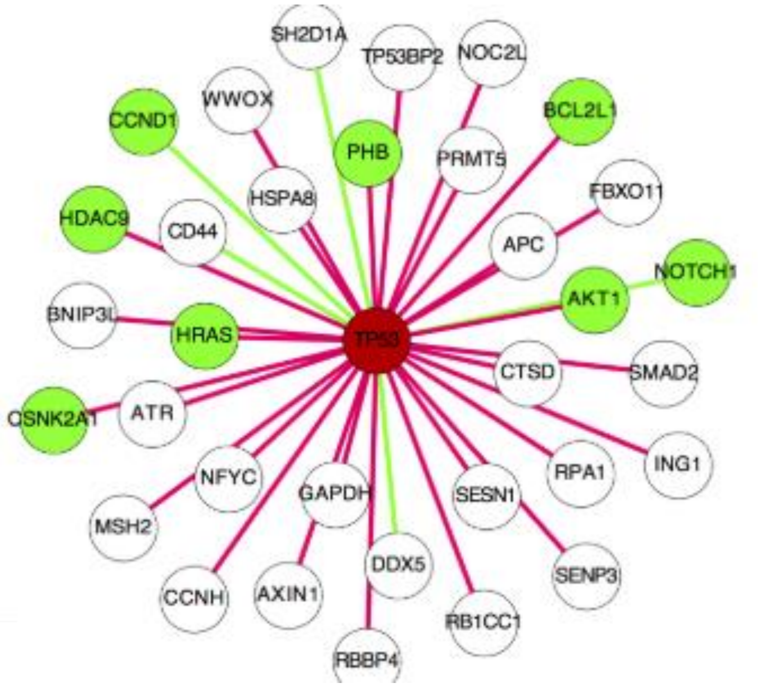
- Perturbation is reductive: different interpretation than the full network.
- Widely applied in many fields (microbiome, transcriptomics, single-cells..)
- Influential paper that inspired further publications.

Sample-specific networks SSN (4/4): results

(Liu et al., 2016)

Interpretation

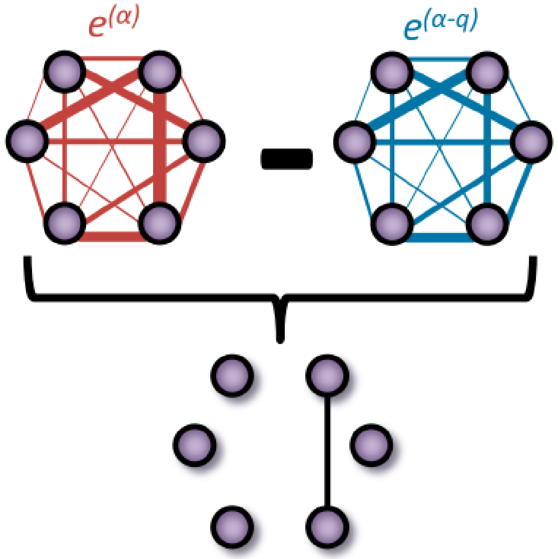
- Individual-specific subnetwork of TP53 (cancer marker) in sample 2574
- Reveal personalized features of **each** sample
- Each cancer type has a specific regulatory pattern
- Initially developed for Pearson correlation – but extendable for every kind of association measure



LIONESS FORMULA (1/3)

Individual network

(Kuijjer et al., 2019)



Sample q's contribution to $e^{(\alpha)}$

- Observation influence

Article

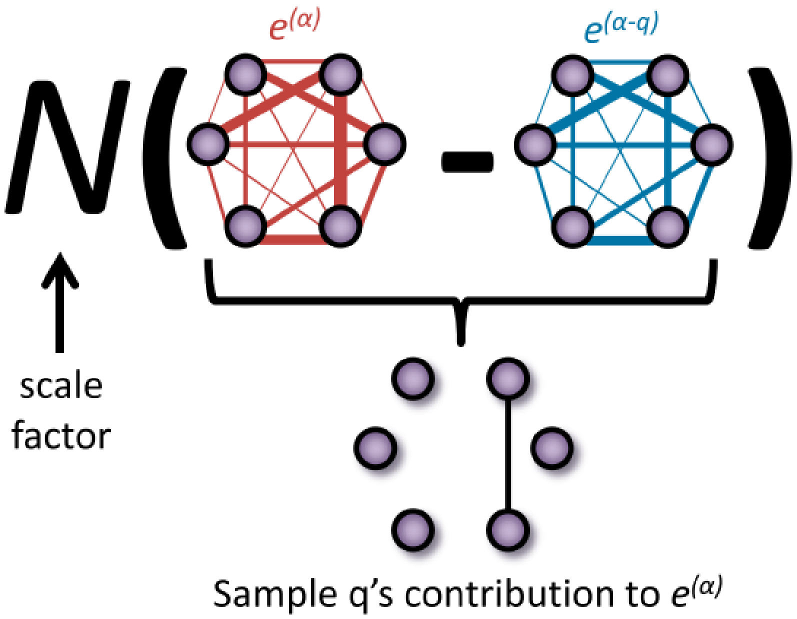
Estimating Sample-Specific Regulatory Networks

Marieke Lydia Kuijjer,[1,7] Matthew George Tung,[2,7] GuoCheng Yuan,[3,4] John Quackenbush,[3,5,6] and Kimberly Glass[5,6,8,*]

LIONESS FORMULA (2/3)

Individual network

(Kuijjer et al., 2019)



- Observation influence
- Scale factor

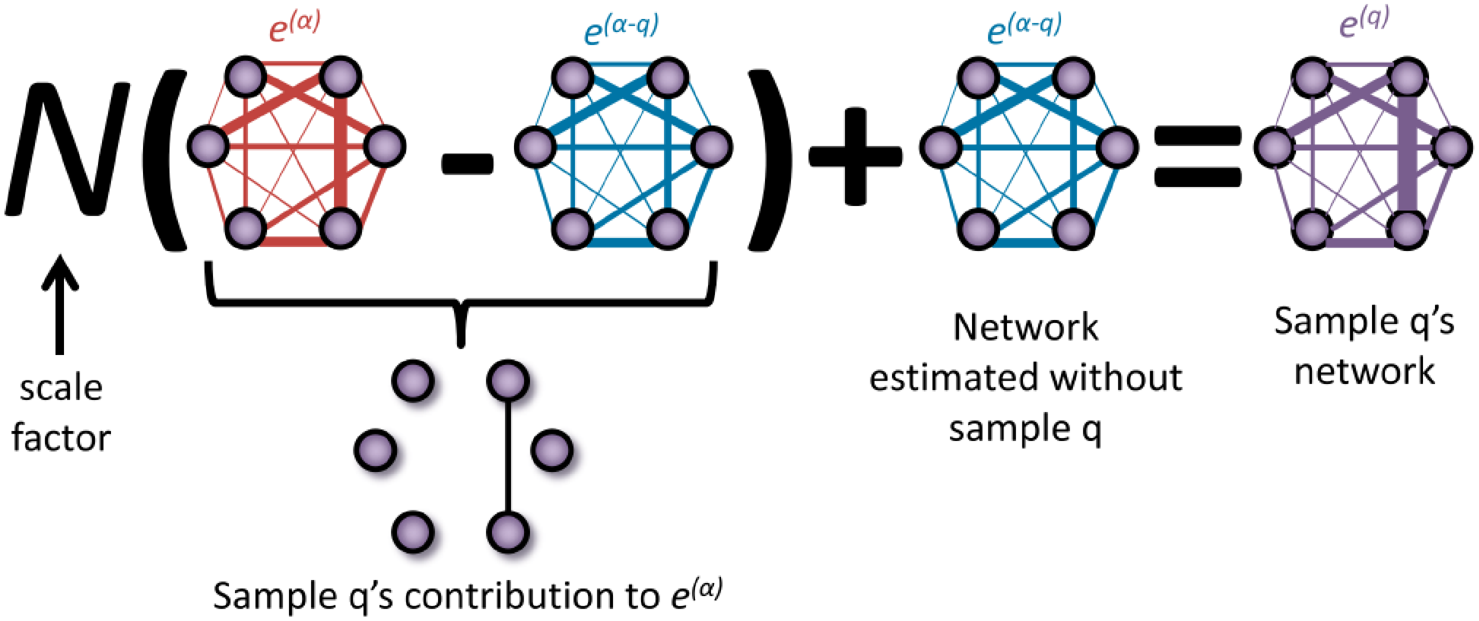
Article

Estimating Sample-Specific Regulatory Networks

Marieke Lydia Kuijjer,[1,7] Matthew George Tung,[2,7] GuoCheng Yuan,[3,4] John Quackenbush,[3,5,6] and Kimberly Glass[5,6,8,*]

LIONESS FORMULA (3/3)

(Kuijjer et al., 2019)



- Observation influence
- Scale factor
- Base network
- N is the number of samples

Article

Estimating Sample-Specific Regulatory Networks

Marieke Lydia Kuijjer,[1,7] Matthew George Tung,[2,7] GuoCheng Yuan,[3,4] John Quackenbush,[3,5,6] and Kimberly Glass[5,6,8,*]

LIONESS - ISNs Comments:

(Kuijjer et al., 2019)

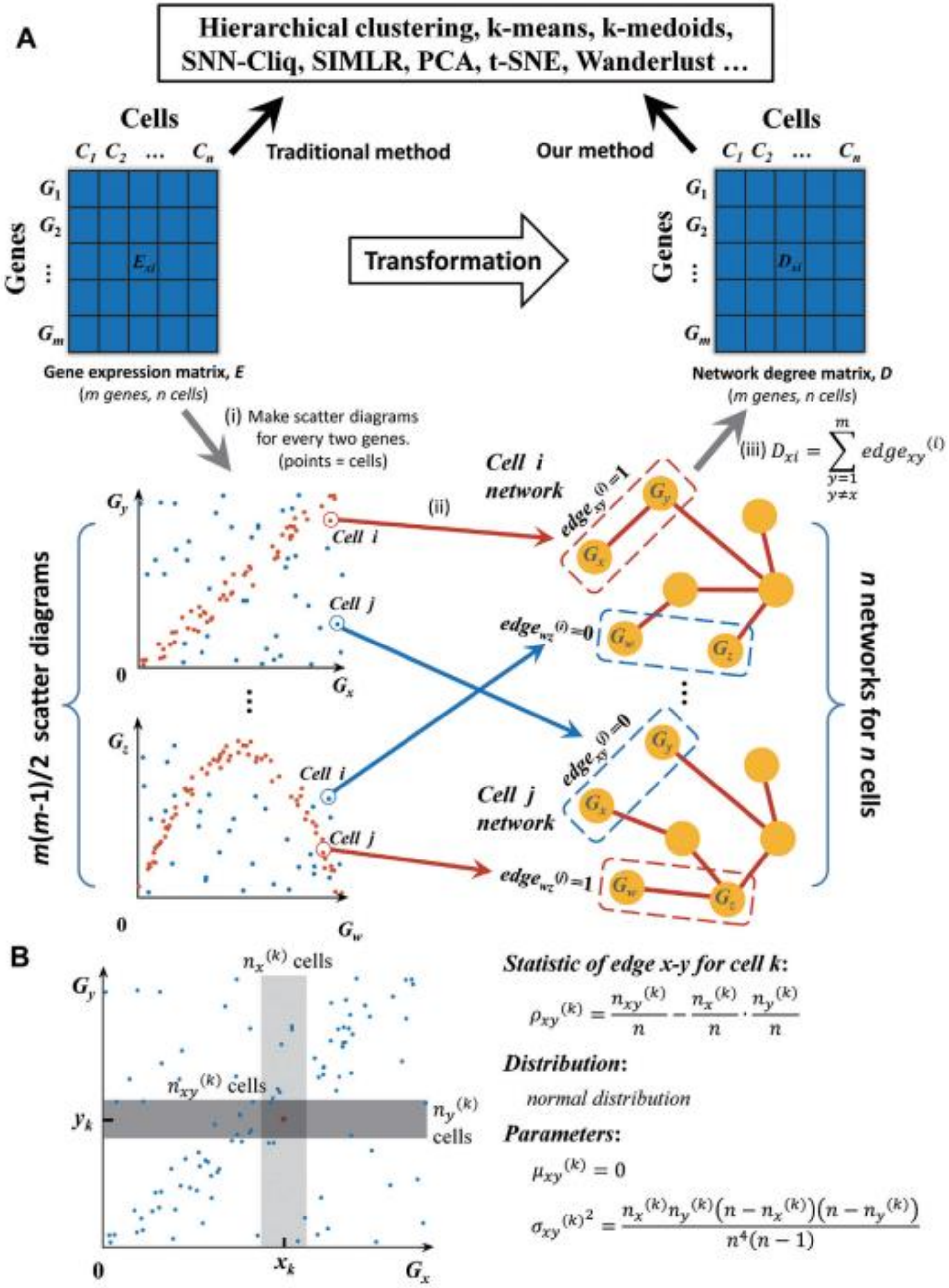
1 How

- No need for reference network: opposite approach than SSN: removal of a sample
- No test statistic for significance:
- Build a **completed** network: same interpretation as the global network

2 Caveats

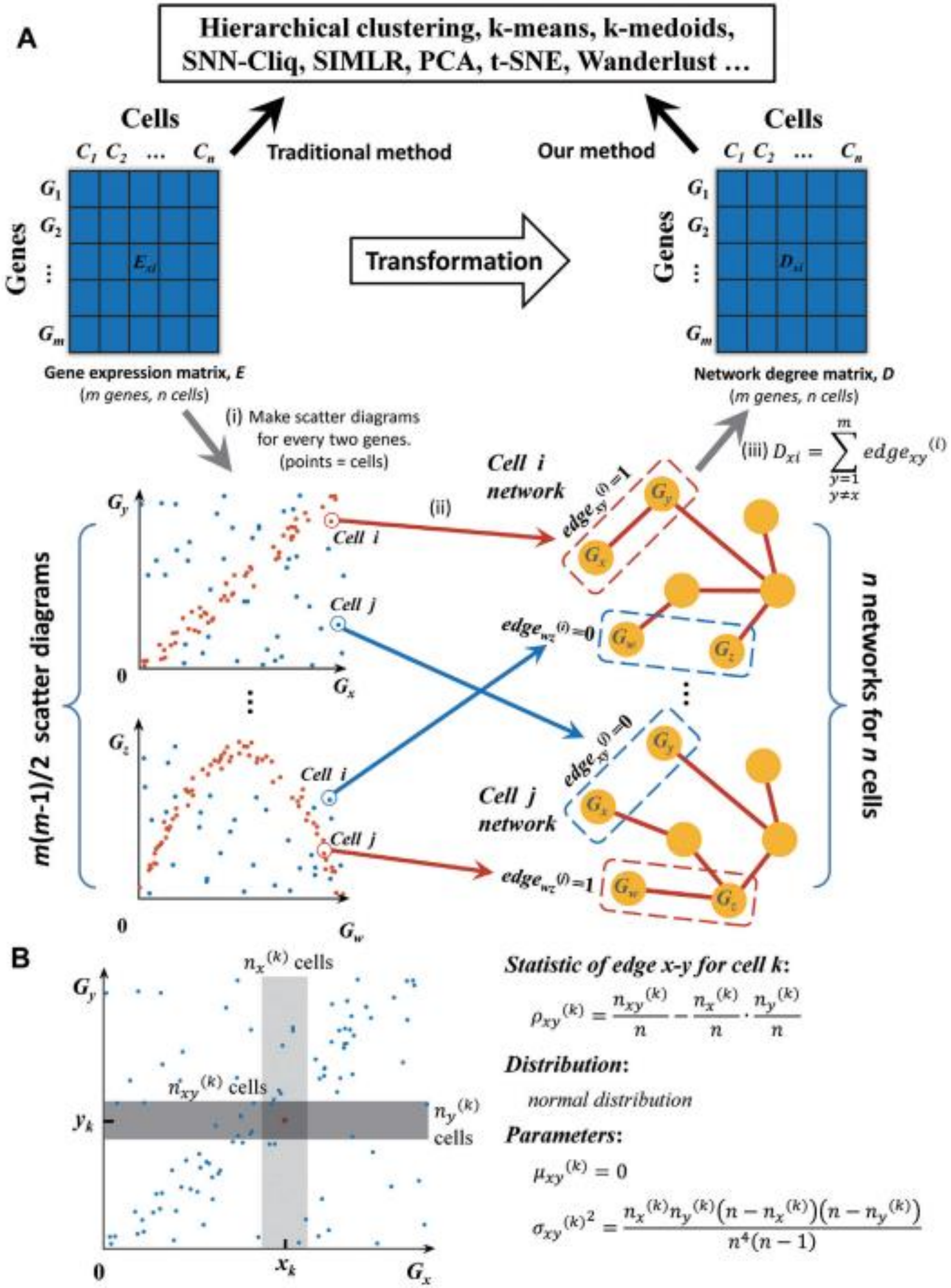
- Time-intensive: the calculation is $O(p^2)$ with p number of nodes-
- Based on the assumptions that the individual-specific edges, on average, represent the aggregate network.
- Not limited from Pearson correlation – it can work with every association mechanism

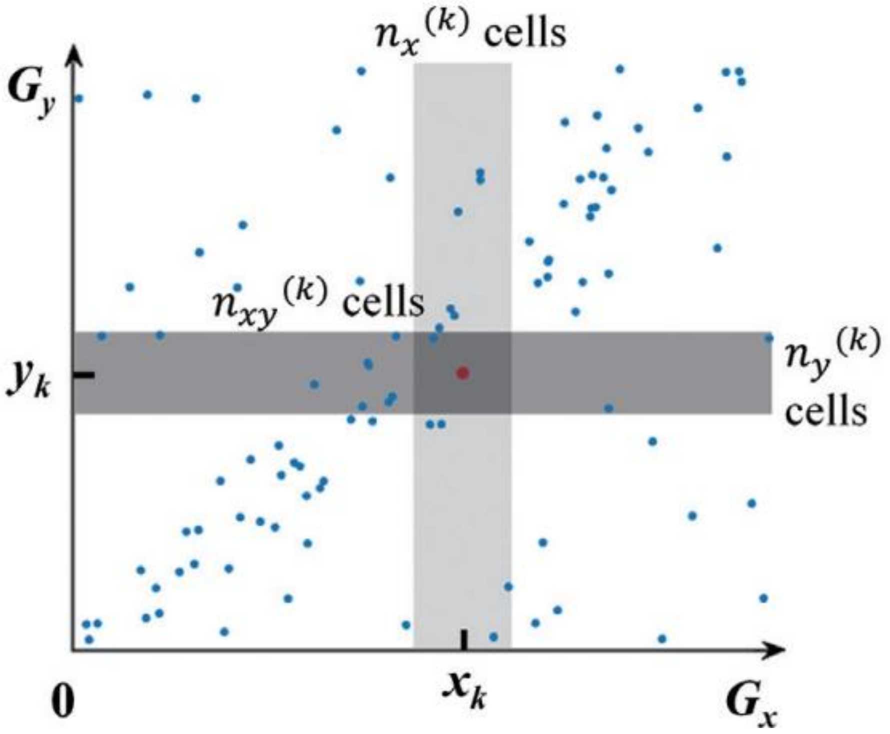
Cell specific network – CSN :


A
Hierarchical clustering, k-means, k-medoids, SNN-Cliq, SIMLR, PCA, t-SNE, Wanderlust …
Traditional method
Our method
Cells
C_1 C_2 … C_n
Genes
G_1 G_2 ⋮ G_m
E_{xi}
Gene expression matrix, E
(m genes, n cells)
Transformation
D_{xi}
Network degree matrix, D
(m genes, n cells)
(i) Make scatter diagrams for every two genes. (points = cells)
(ii)
(iii) $D_{xi} = \sum_{y=1, y\neq x}^{m} edge_{xy}^{(i)}$
Cell i network
Cell j network
$edge_{xy}^{(i)}=1$
$edge_{wz}^{(i)}=0$
$edge_{xy}^{(j)}=0$
$edge_{wz}^{(j)}=1$
G_x G_y G_w G_z
Cell i
Cell j
0
m(m-1)/2 scatter diagrams
n networks for n cells
B
$n_x^{(k)}$ cells
$n_{xy}^{(k)}$ cells
$n_y^{(k)}$ cells
y_k
x_k
Statistic of edge x-y for cell k:
$\rho_{xy}^{(k)} = \frac{n_{xy}^{(k)}}{n} - \frac{n_x^{(k)}}{n} \cdot \frac{n_y^{(k)}}{n}$
Distribution:
normal distribution
Parameters:
$\mu_{xy}^{(k)} = 0$
$\sigma_{xy}^{(k)2} = \frac{n_x^{(k)} n_y^{(k)} (n - n_x^{(k)})(n - n_y^{(k)})}{n^4 (n-1)}$

(Dai et al., 2019)

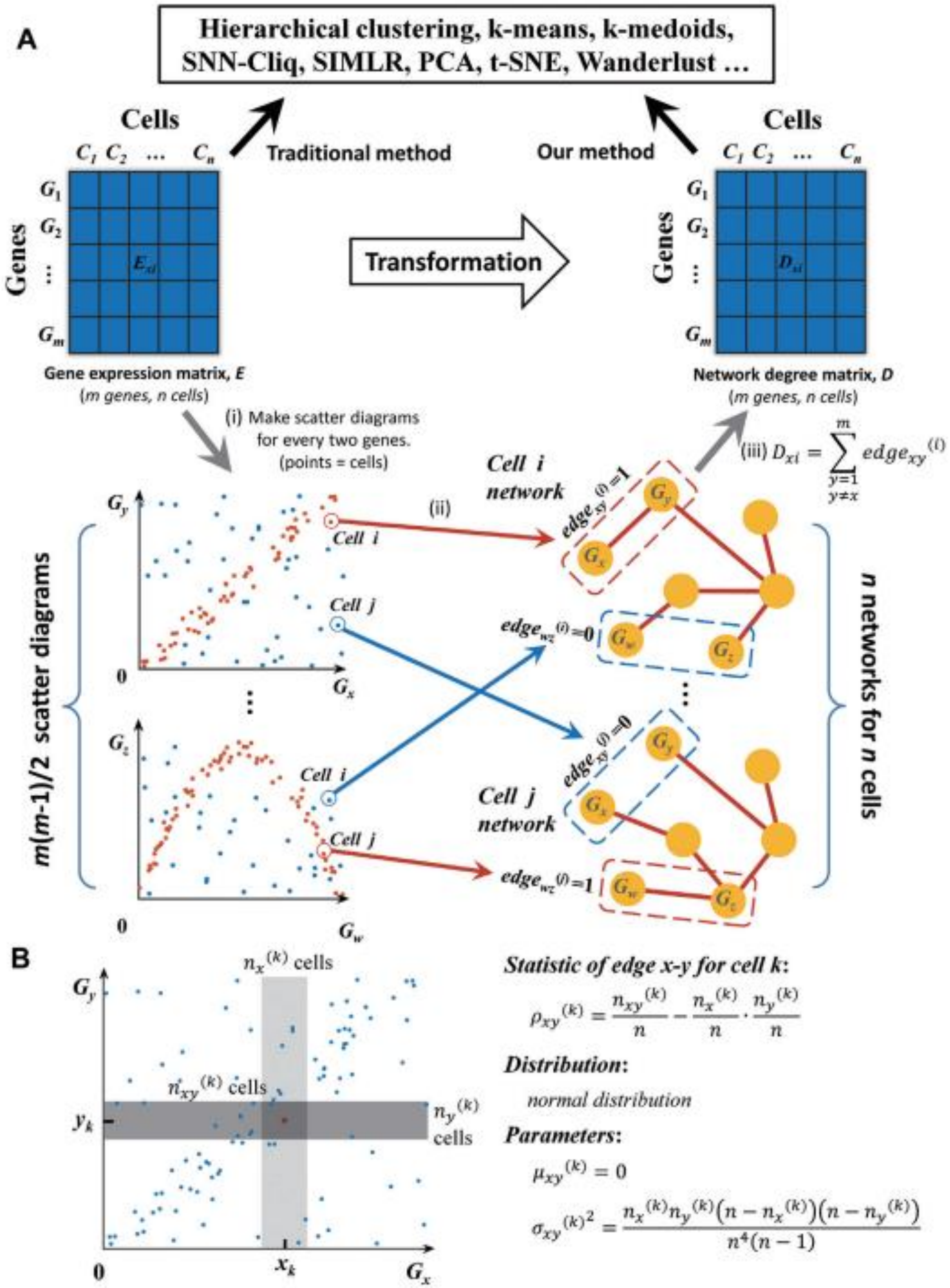
Cell specific network – CSN :

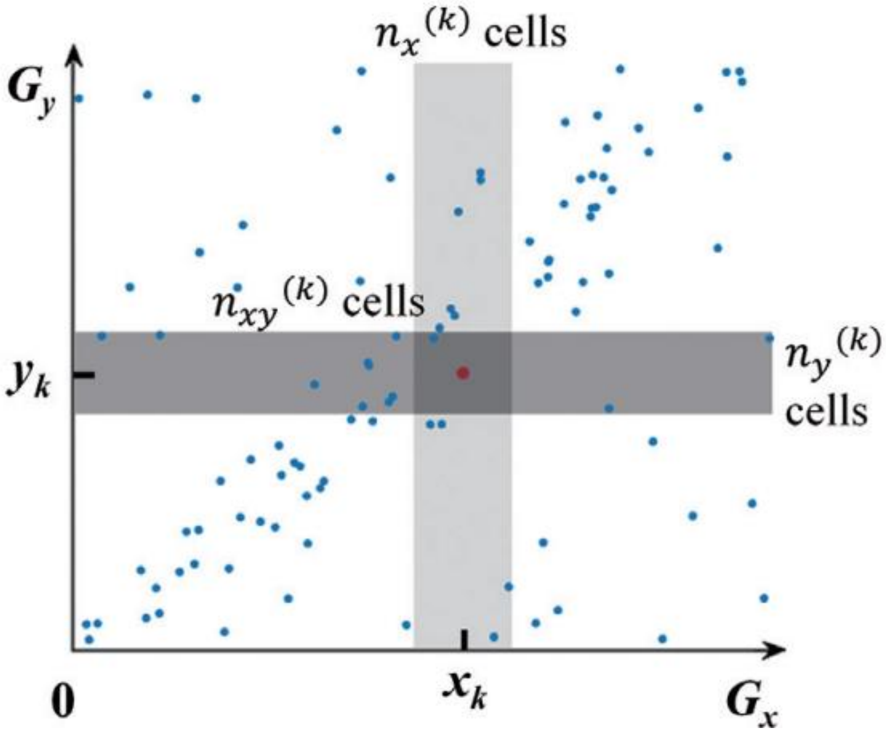

A
Hierarchical clustering, k-means, k-medoids, SNN-Cliq, SIMLR, PCA, t-SNE, Wanderlust …
Traditional method
Our method
Cells
C_1 C_2 … C_n
Genes
G_1 G_2 ⋮ G_m
E_{xi}
Gene expression matrix, E
(m genes, n cells)
Transformation
D_{xi}
Network degree matrix, D
(m genes, n cells)
(i) Make scatter diagrams for every two genes. (points = cells)
(ii)
(iii) $D_{xi} = \sum_{y=1, y\neq x}^{m} edge_{xy}^{(i)}$
m(m-1)/2 scatter diagrams
G_y G_x G_z G_w 0
Cell i
Cell j
Cell i network
Cell j network
$edge_{xy}^{(i)}=1$
$edge_{wz}^{(i)}=0$
$edge_{xy}^{(j)}=0$
$edge_{wz}^{(j)}=1$
n networks for n cells
B
$n_x^{(k)}$ cells
$n_{xy}^{(k)}$ cells
$n_y^{(k)}$ cells
y_k x_k
Statistic of edge x-y for cell k:
$\rho_{xy}^{(k)} = \frac{n_{xy}^{(k)}}{n} - \frac{n_x^{(k)}}{n} \cdot \frac{n_y^{(k)}}{n}$
Distribution:
normal distribution
Parameters:
$\mu_{xy}^{(k)} = 0$
$\sigma_{xy}^{(k)2} = \frac{n_x^{(k)} n_y^{(k)} (n - n_x^{(k)})(n - n_y^{(k)})}{n^4(n-1)}$



(Dai et al., 2019)

Cell specific network – CSN :





***Statistic of edge x-y for cell k*:**

$$\rho_{xy}^{(k)} = \frac{n_{xy}^{(k)}}{n} - \frac{n_x^{(k)}}{n}\cdot\frac{n_y^{(k)}}{n}$$

***Distribution*:**

normal distribution

***Parameters*:**

$$\mu_{xy}^{(k)} = 0$$

$$\sigma_{xy}^{(k)^2} = \frac{n_x^{(k)}n_y^{(k)}(n-n_x^{(k)})(n-n_y^{(k)})}{n^4(n-1)}$$

Individual network

Cell specific network – (C)CSN :

(Li et al., 2021)

1 How

- Starting from the single cell(sample) value for a gene pair (x, y), we depict a 10% **UNIVARIATE** interval of all samples and compute how many obs are into the intersection n_{xy}.
- If ${}^{n_{xy}}/_{n} = {}^{n_x}/_{n} * {}^{n_y}/_{n}$, i.e. does not refuse independence hypothesis: no edge, if it is up or down regulated
- If p-value < 1%: edge
- Build a **binary** network

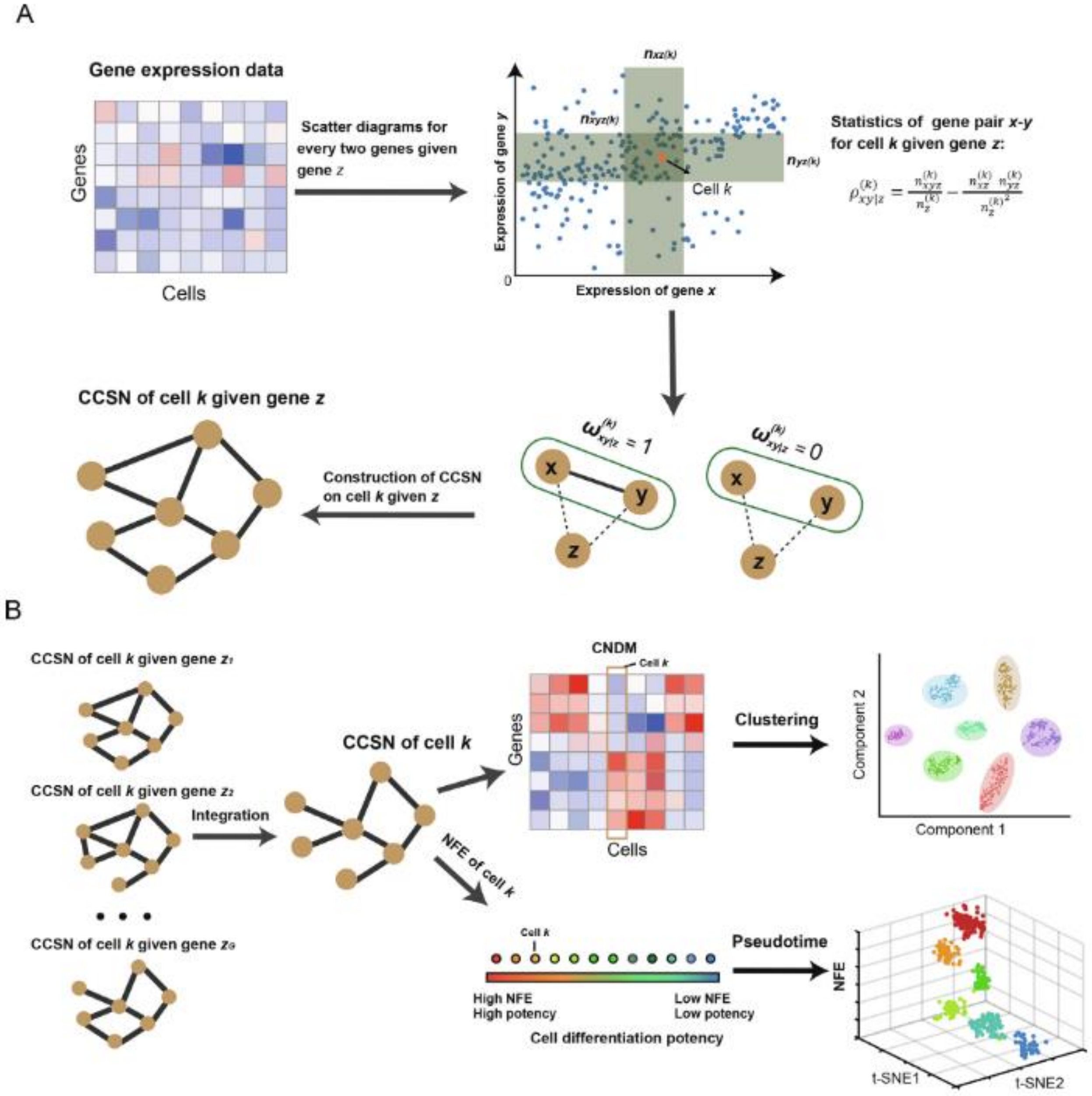
2 Caveats

- No reference network
- Density-based

Cell specific network – CCSN :

Individual network

(Li et al., 2021)


A
Gene expression data
Genes
Cells
Scatter diagrams for every two genes given gene z
$n_{xz(k)}$
$n_{xyz(k)}$
$n_{yz(k)}$
Cell k
Expression of gene y
Expression of gene x
0
Statistics of gene pair x-y for cell k given gene z:
$\rho^{(k)}_{xy|z} = \frac{n^{(k)}_{xyz}}{n^{(k)}_{z}} - \frac{n^{(k)}_{xz}\, n^{(k)}_{yz}}{n^{(k)\,2}_{z}}$
CCSN of cell k given gene z
Construction of CCSN on cell k given z
$\omega^{(k)}_{xy|z} = 1$
$\omega^{(k)}_{xy|z} = 0$
x
y
z
B
CCSN of cell k given gene z_1
CCSN of cell k given gene z_2
CCSN of cell k given gene z_G
Integration
CCSN of cell k
CNDM
Cell k
Genes
Cells
Clustering
Component 2
Component 1
NFE of cell k
Cell k
High NFE
High potency
Low NFE
Low potency
Cell differentiation potency
Pseudotime
NFE
t-SNE1
t-SNE2

Cell specific network – CCSN: formula

Individual network

(Li et al., 2021)

- For a sample (cell), the statistic is

$$p_{x,y|z} = \frac{n_{xyz}}{n} - \frac{n_{xz}n_{yz}}{n^2}$$

- Representing the probability (with the 10% threshold) of having values x,y,z of genes X, Y, Z
- Difference from this observed probability to the ones if X and Y were independent
- Normalization with expected value and standard deviation:

$$\mu_{xy|z} = 0 \qquad \sigma_{xy|z} = \sqrt{\frac{n_{xz}n_{yz}(n_z - n_{xz})(n_z - n_{yz})}{n_z(n_z - 1)}}$$

- Normalized statistic:

$$\hat{p}_{xy|z} = \frac{p_{x,y|z} - \mu_{xy|z}}{\sigma_{xy|z}}$$

Individual network

Cell specific network – CCSN: pipeline

(Li et al., 2021)

- For a sample (cell), 1st construct a CSN without conditional genes: the edge between gene x and y are determined with CSN:

$$edge_{x,y} = \begin{cases} 1 & genes\ x\ and\ y\ are\ dependent \\ \overline{0} & \overline{genes\ x\ andy\ are\ indipendent} \end{cases}$$

- We calculate the node degree as the importance of the node

$$D_z = \sum_{y=1, y\neq z}^{M} edge_{zy}$$

- Top G largest importance genes as the conditional genes

$$\{z_g, g = 1,2,3, ..G\} \rightarrow \{C_{z1}, \ldots, C_{zG}\}$$

- Calculate CCSN based on the conditional gene set:
- Hence, we have G CCSN for each sample
- Merge those into the final CCSN

$$\overline{C_k} = \frac{1}{G}\sum_{g=1}^{g} C_{zg}$$

Cell specific network – CCSN :

(Li et al., 2021)

1 How

- Same structure as CSN – but added conditionality.
- Iterative procedure of estimating CSN – finding the driving nodes – and use those for CCSN
- Parameters: width of the univariate interval for samples; cut-off to determine which are the driving nodes, threshold for significance.

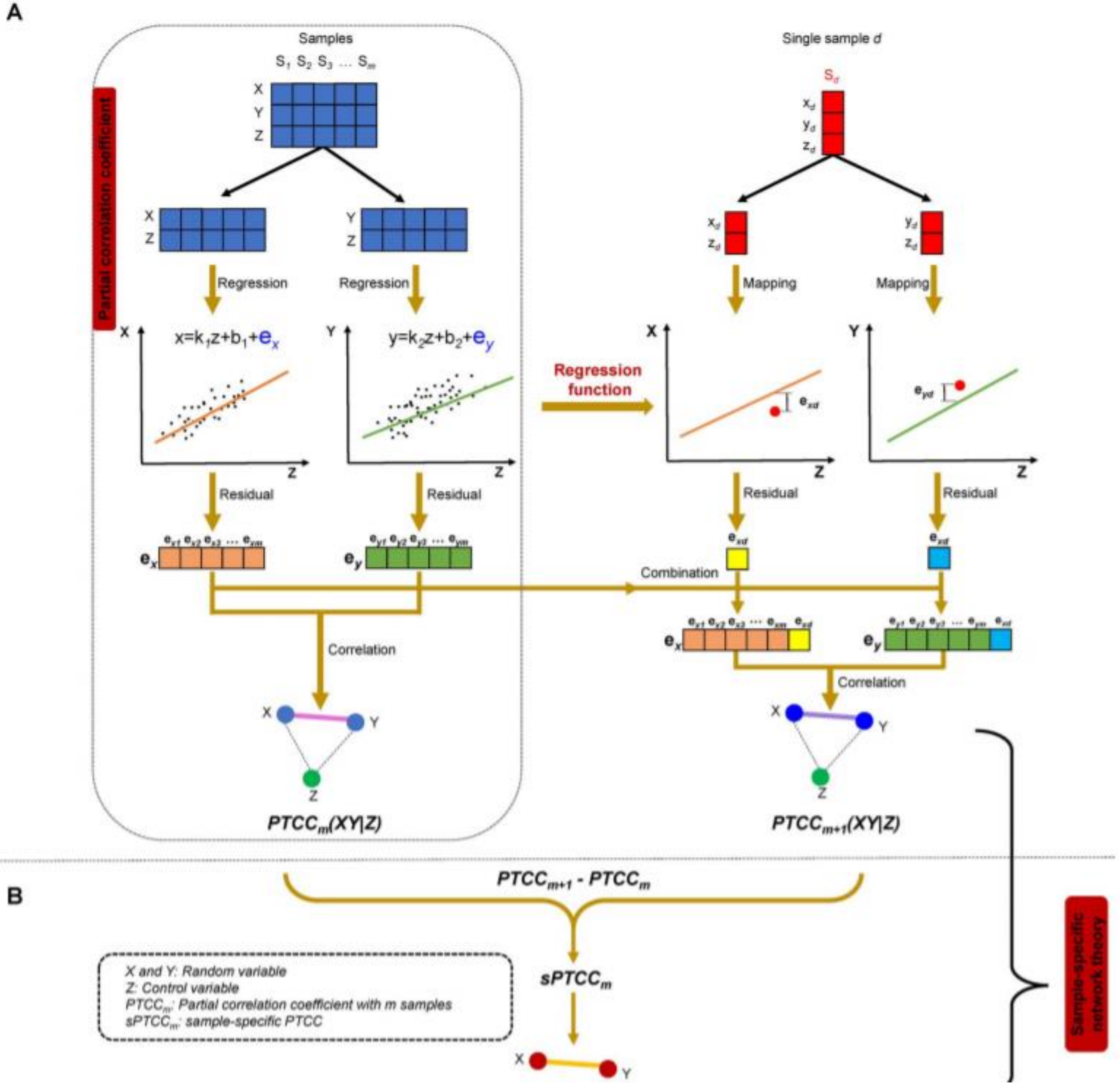
2 Caveats

- Parameter-dependent: questionable stability
- Used in single-cells
- 2-step procedure

Individual network

Partial network P-SSN

(Huang et al., 2021)



Partial network P-SSN

(Huang Y et al. 2021)

1 How

- Partial correlation (on PCC) with considering a variable Z
- After creating a background network with controls, for a pair X and Y, Z is considered "confounder" and to take into account if its correlation with both X and Y is > 0.7
- Gene pairs retained in the global network are the X,Y that have significant p-value (with a T-test, 0.01) with ALL possible variable Z that satisfy the condition before.
- Then, a sample is added and the sPTCC is calculated.
- Using Liu et al., significant sPTCC (p-value < 0.05) with ALL possible Z confounders

Partial network P-SSN

(Huang Y et al. 2021)

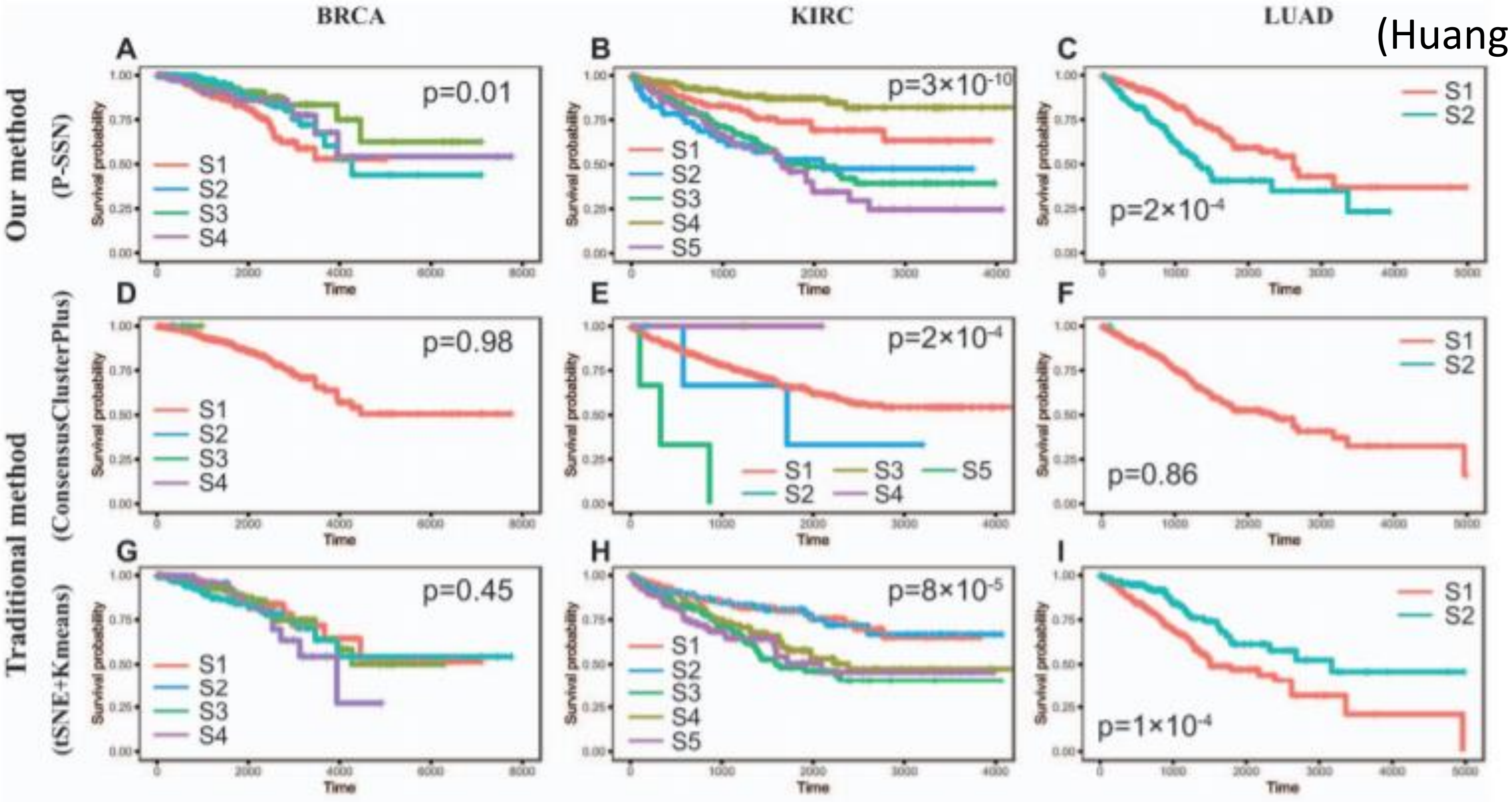
2 Caveats

- Based on Pearson correlation
- Perturbation network: does not reconstruct full network
- Parameters: 0.7 for “high correlation”; threshold for T-test p-value
- Using regression’s residuals

Partial network P-SSN

Individual network

(Huang Y et al. 2021)



J

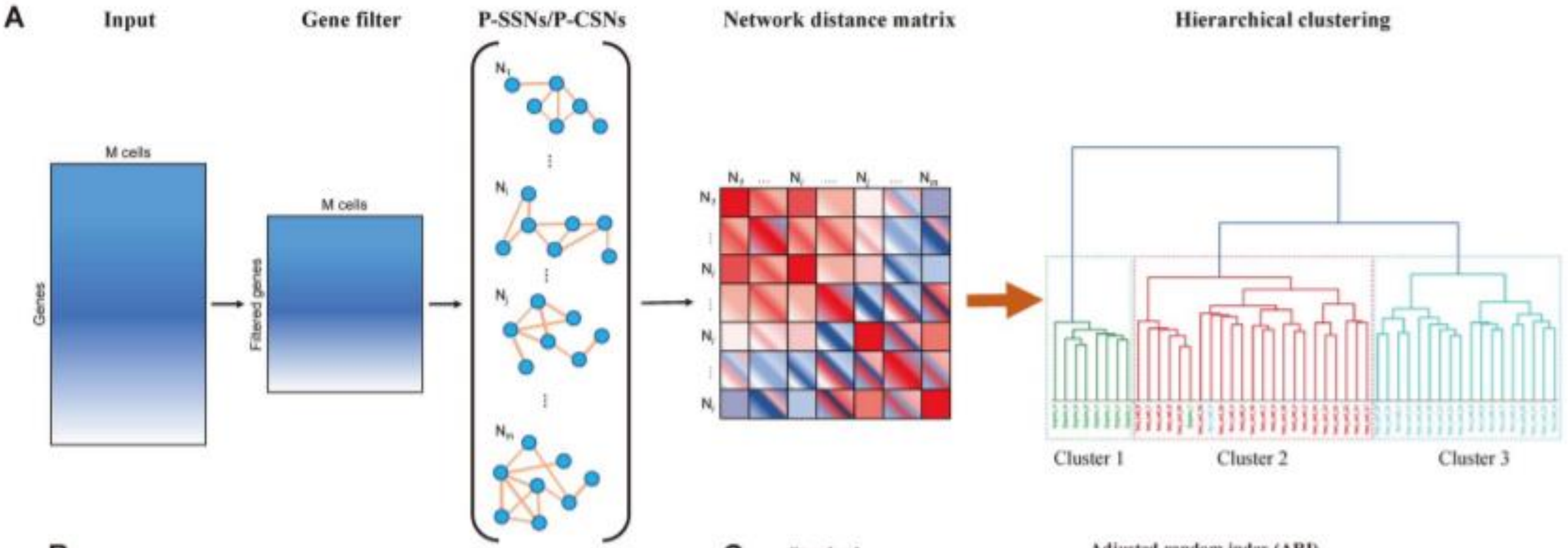
Cancer	**Network**	Euclidean	Cosine	Correlation	Braycurtis	Canberra	Chebyshev	Kulsinski	Sqeuclidean	Jaccard
BRCA	**0.01**	0.93	0.96	0.98	0.03	0.04	0.7	0.013	0.75	0.92
KIRC	**3×10^{-10}**	0.012	1×10^{-8}	2×10^{-6}	4×10^{-10}	2×10^{-9}	0.012	0.08	0.72	0.23
LUAD	**2×10^{-4}**	0.013	0.002	0.001	0.56	0.03	0.013	0.01	0.07	0.78

Figure 4. The survival curves for subtyping three cancers. (A) P-SSN method for BRCA. (B) P-SSN method for KIRC. (C) P-SSN method for LUAD. (D) ConsensusClusterPlus for BRCA. (E) ConsensusClusterPlus for KIRC. (F) ConsensusClusterPlus for LUAD. (G) tSNE + Kmeans for BRCA. (H) tSNE + Kmeans for KIRC. (I) tSNE + Kmeans for LUAD. (J) Comparison between network distance and other nine traditional distances in the subtype identification for BRCA, KIRC and LUAD. The figure showed the log-rank P-value of survival analysis for the subtypes of three tumors, and the subtypes were obtained by hierarchical clustering algorithm based on different distances. The bold values were the best results in every row.

Partial network P-SSN

Individual network

(Huang Y et al. 2021)



B

Clustering	LUAD	KIRC	BRCA
	Log-rank P-value		
P-SSN	**2×10^{-4}**	**3×10^{-11}**	**0.01**
ConsensusClusterPlus	0.86	0.0002	0.98
tSNE+Kmeans	0.0001	0.00008	0.45

C

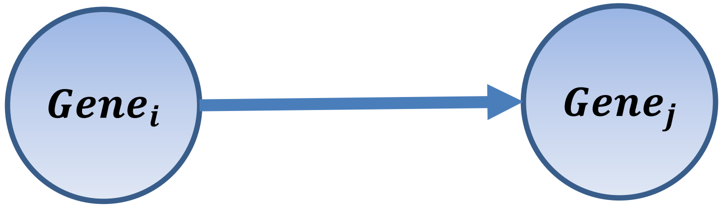
Standard	Adjusted random index (ARI)					
Dataset	**P-CSN**	SEURAT	SNN-Cliq	SINCERA	tSNE+Kmeans	pcaReduce
Basic	**0.87**	0.00	0.62	0.29	0.13	0.59
Yan	**0.86**	0.00	0.52	0.35	0.54	0.68
Goolam	**0.43**	0.03	0.17	0.25	0.52	0.43
Deng	**0.51**	0.42	0.35	0.43	0.46	0.45
Ting	**0.72**	0.05	0.39	0.32	0.65	0.39
Treutlein	**0.32**	0.00	0.23	0.28	0.20	0.31
Pollen	**0.89**	0.69	0.43	0.53	0.65	0.48
Kim	**0.61**	0.02	0.58	0.78	0.37	0.20

Figure 6. The P-SSN/P-CSN clustering based on network distance. (A) The framework of P-SSN/P-CSN clustering based on network distance. (B) The comparison between P-SSN clustering, ConsensusClusterPlus, and tSNE + Kmeans in subtypes identification for LUAD, KIRC and BRCA, evaluated by the log-rank P-value of survival analysis. (C) The comparison between P-CSN and SEURATE, SNN-Clip, SINCERA, tSNE+kmeans, pcaReduce in clustering of scRNA-seq data, evaluated by ARI.

Direct network

Interpretation

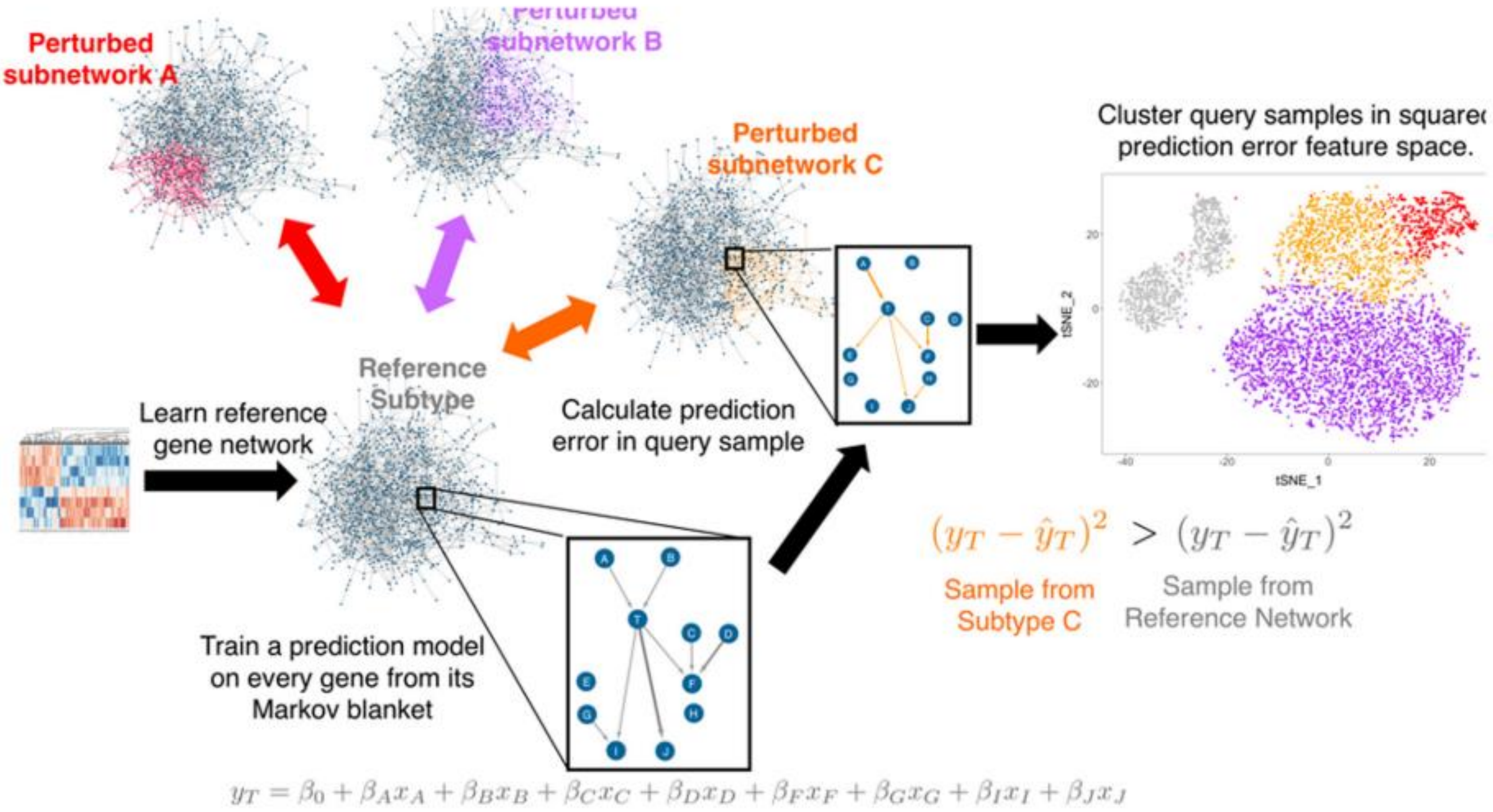
- Directed edges originate from causal mechanism



Direct network: ssNPA

Individual network

(Buschur KL et al. 2020)


Perturbed subnetwork A
Perturbed subnetwork B
Perturbed subnetwork C
Cluster query samples in squared prediction error feature space.
tSNE_2
tSNE_1
20
0
-20
-40
Reference Subtype
Learn reference gene network
Calculate prediction error in query sample
Train a prediction model on every gene from its Markov blanket
A
B
T
C
D
E
F
G
H
I
J
$(y_T - \hat{y}_T)^2 > (y_T - \hat{y}_T)^2$
Sample from Subtype C
Sample from Reference Network
$y_T = \beta_0 + \beta_A x_A + \beta_B x_B + \beta_C x_C + \beta_D x_D + \beta_F x_F + \beta_G x_G + \beta_I x_I + \beta_J x_J$

Direct network: ssNPA

(Buschur KL et al. 2020)

1 How

- Build a reference network with control only
- Add a case sample

ssNPA build a predictive model for every gene based on the Markov blanker

- Applied to a new sgene, for a case, produce a prediction
- Residuals predicted – real value = residuals, one for each gene
- Residuals used to
  - Cluster samples (genes into groups
  - Assess group characteristic
  - Assign individual patient into a disease subgroup

Individual network

Direct networks ssNPA

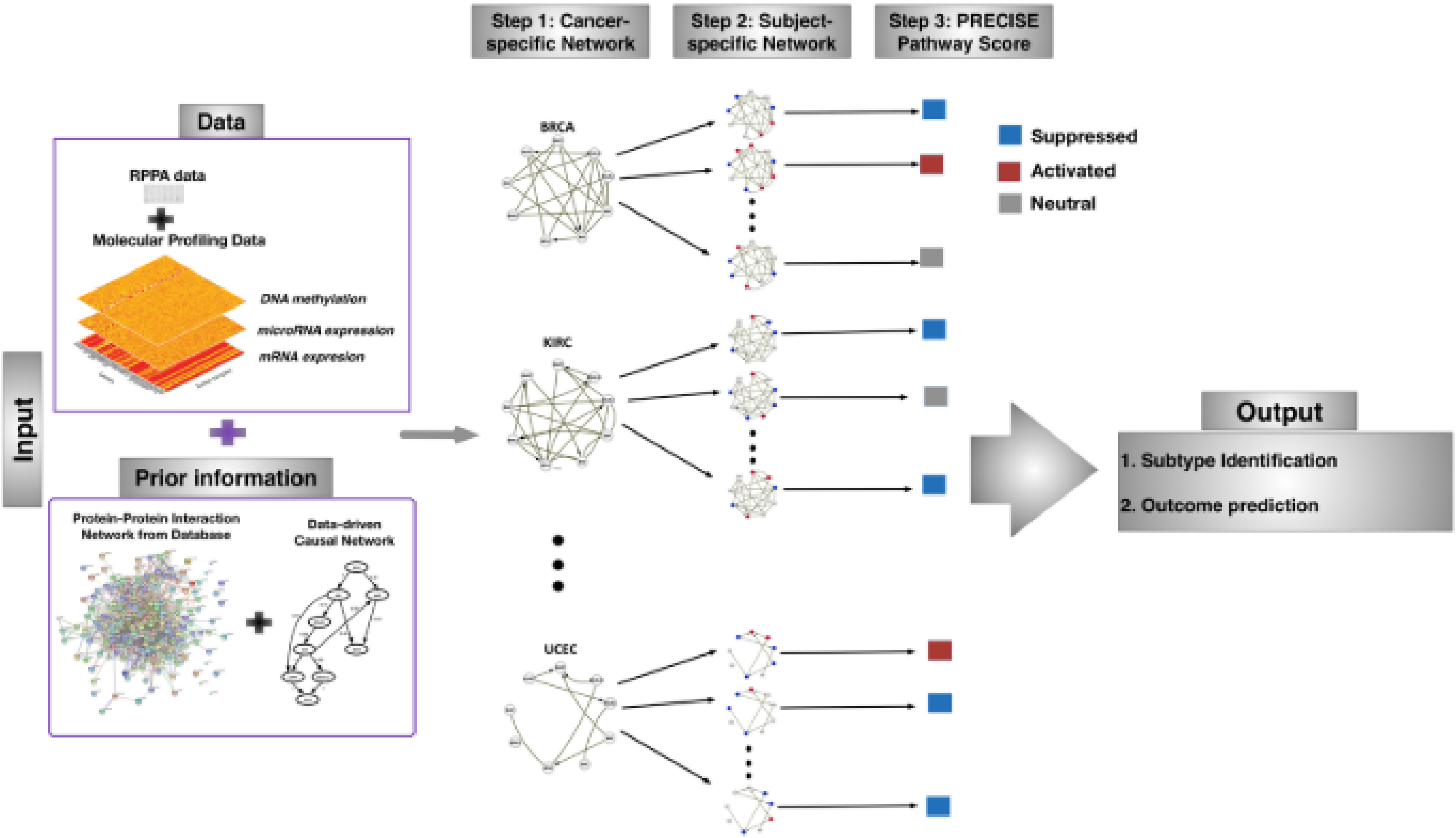
(Buschur KL et al. 2020)

2 Caveats

- Not an ISN, create a global network
- Residual-based
- Directed network
- Use of a Markov blanket

Directed networks PRECISE

(Ha et al., 2018)



Step 1: Cancer-specific Network

Step 2: Subject-specific Network

Step 3: PRECISE Pathway Score

Input

Data

Prior information

Output

1. Subtype Identification
2. Outcome prediction

🟦 Suppressed

🟥 Activated

⬜ Neutral

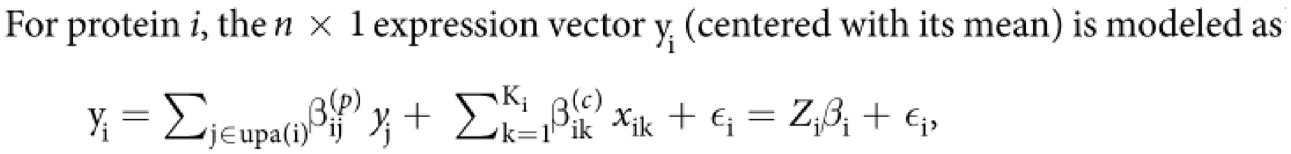
Directed networks PRECISE

Individual network

(Ha et al., 2018)

1 How

- PPI causal network estimated and combined with prior information
- Bayesian estimation of integrated cancer-specific networks
- W_{ij} weight for protein $i \rightarrow j$, if i regulator of j
  - Decided with prior inclusion information
  - $W_{ij} \neq W_{ji}$



For protein i, the $n \times 1$ expression vector y_i (centered with its mean) is modeled as

$$y_i = \sum\nolimits_{j \in \mathrm{upa}(i)} \beta_{ij}^{(p)} y_j + \sum\nolimits_{k=1}^{K_i} \beta_{ik}^{(c)} x_{ik} + \epsilon_i = Z_i \beta_i + \epsilon_i,$$

- Select Posterior probability > 0.5
- PRECISE network: patient specific labels.
- Network structure is fixed, only the label change

Directed networks PRECISE

(Ha et al., 2018)

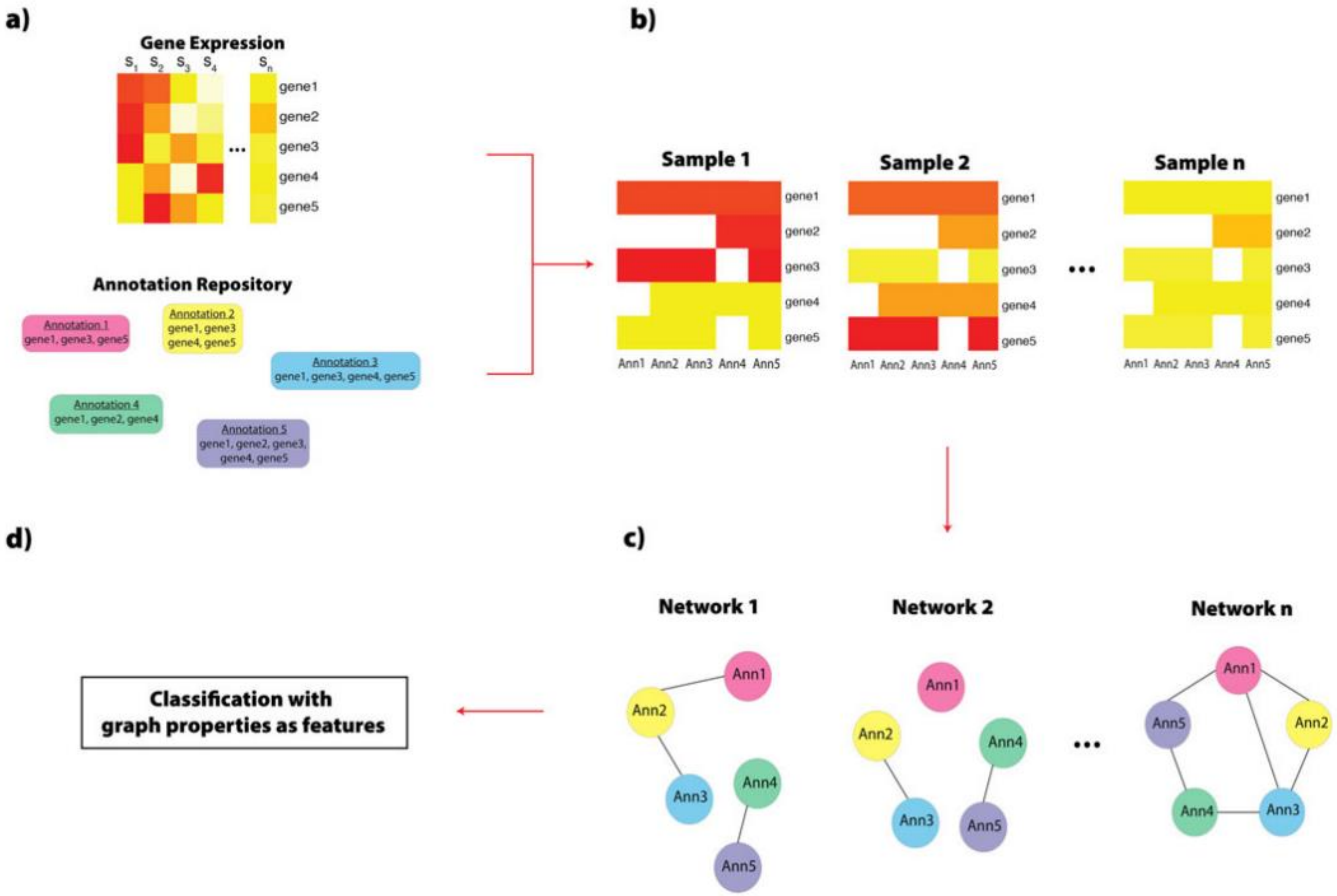
2 Caveats

- Individual-specificity on the nodes.
- Directed network
- Personalize label of cancer-specific networks
- Use of a Markov blanket

Other: PAN Pipeline

Individual network

(Nguyen et al., 2021)


a)
Gene Expression
S_1 S_2 S_3 S_4 ... S_n
gene1
gene2
gene3
gene4
gene5
Annotation Repository
Annotation 1 gene1, gene3, gene5
Annotation 2 gene1, gene3 gene4, gene5
Annotation 3 gene1, gene3, gene4, gene5
Annotation 4 gene1, gene2, gene4
Annotation 5 gene1, gene2, gene3, gene4, gene5
b)
Sample 1
Sample 2
Sample n
gene1
gene2
gene3
gene4
gene5
Ann1 Ann2 Ann3 Ann4 Ann5
...
c)
Network 1
Network 2
Network n
Ann1
Ann2
Ann3
Ann4
Ann5
...
d)
Classification with graph properties as features

Individual network

Other: PAN Pipeline

(Nguyen et al., 2021)

1 How

- Use annotations: known biological connections
- Cluster annotations and use similarities as edge: Euclidean distance (0 if same sets of gene); selected top edges (smallest distance)
- Calculate graph statistics: closeness centrality; betweenness centrality; PageRank; Use them to predict Relapse/Non relapse

2 Caveats

- Arbitrary choice of «top edges»
- Weak univariate performances
- Heavily dependent on the annotations

Individual network

Other: PAN Results

(Nguyen et al., 2021)

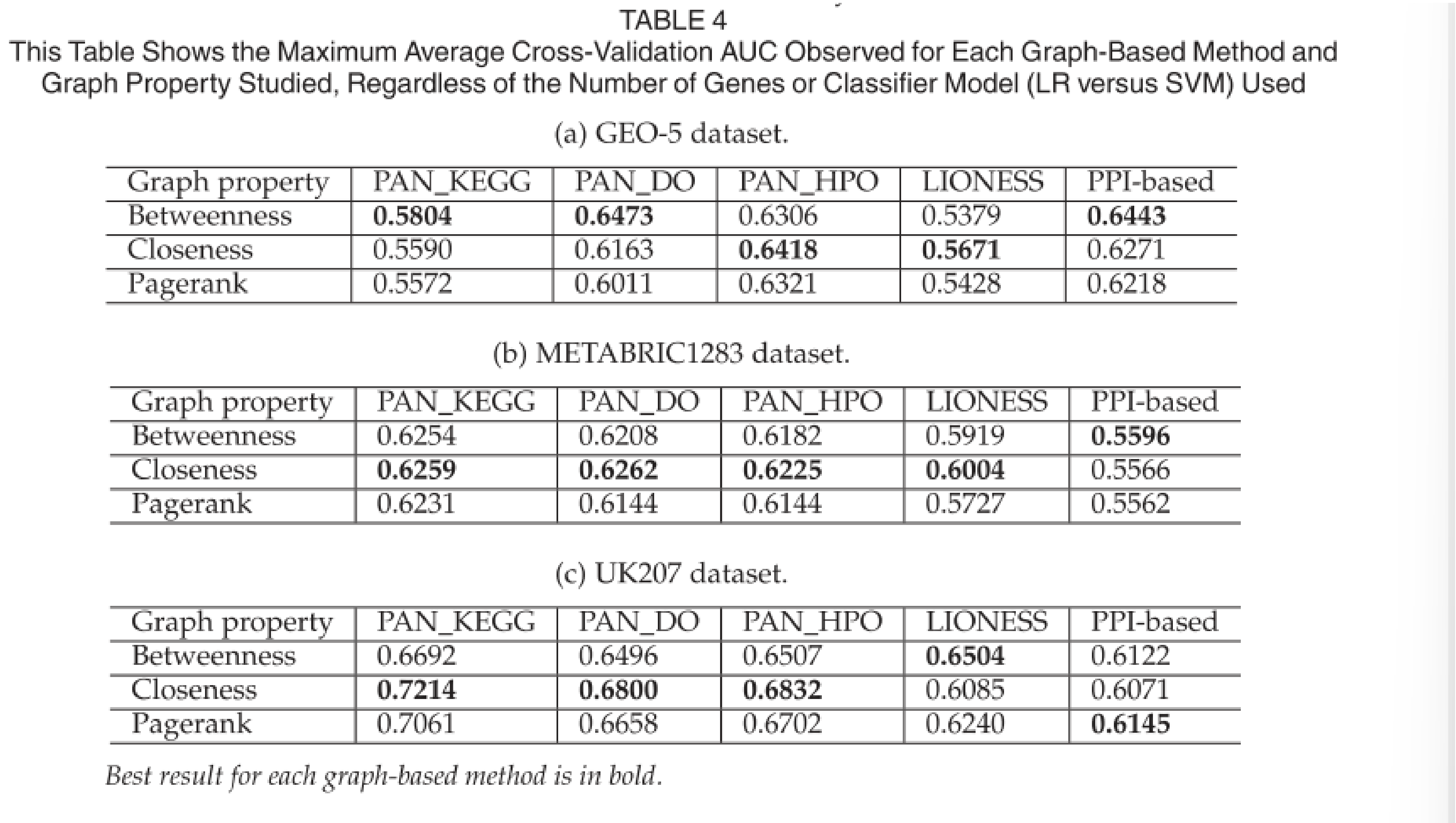


TABLE 4
This Table Shows the Maximum Average Cross-Validation AUC Observed for Each Graph-Based Method and Graph Property Studied, Regardless of the Number of Genes or Classifier Model (LR versus SVM) Used

(a) GEO-5 dataset.

Graph property	PAN_KEGG	PAN_DO	PAN_HPO	LIONESS	PPI-based
Betweenness	**0.5804**	**0.6473**	0.6306	0.5379	**0.6443**
Closeness	0.5590	0.6163	**0.6418**	**0.5671**	0.6271
Pagerank	0.5572	0.6011	0.6321	0.5428	0.6218

(b) METABRIC1283 dataset.

Graph property	PAN_KEGG	PAN_DO	PAN_HPO	LIONESS	PPI-based
Betweenness	0.6254	0.6208	0.6182	0.5919	**0.5596**
Closeness	**0.6259**	**0.6262**	**0.6225**	**0.6004**	0.5566
Pagerank	0.6231	0.6144	0.6144	0.5727	0.5562

(c) UK207 dataset.

Graph property	PAN_KEGG	PAN_DO	PAN_HPO	LIONESS	PPI-based
Betweenness	0.6692	0.6496	0.6507	**0.6504**	0.6122
Closeness	**0.7214**	**0.6800**	**0.6832**	0.6085	0.6071
Pagerank	0.7061	0.6658	0.6702	0.6240	**0.6145**

Best result for each graph-based method is in bold.

Features:	SSN	LIONESS	CSN	C-CSN	P-SSN	PAN	ssNPA	PRECISE
Nodes	Genes	Genes	Genes / single cell	Genes / single cell	Genes	Annotation	Genes	Genes
Type of network	Perturbation	Completed	NaN	NaN	Perturbed	NaN	Perturbed	Completed
Directionality	Undirected	Undirected	Undirected	Undirected	Undirected	Undirected	Directed	Directed
Confounders ?	No	No	No	Yes, considers the partial correlation to driver genes	Yes, considers the partial correlation to genes associated with both	No	N	Yes, consider external covariate
Reference network	Needed, built with control samples	Not needed	Not needed	Not needed	Needed, built with control samples	Not needed	Yes, control only	No
Individual-specific	Y	Y	Y	Y	Y	Y	N	No IS-edges, only IS-nodes
Parameters	Significance threshold: usually 5%	No parameter: only type of association	Span of univariate interval: 10% Significance threshold: 1-5%	Span of univariate interval: 10% Significance threshold: 1-5%; # top driver genes	0.7 for high correlation gene selection; Significance threshold: 1-5%; Gene pairs significance	#Top edges	PD parameter for FGES	Posterior probability of inclusion = 0.5; Alpha = 0.01 for prior inclusion probability
Type of association	Pearson correlation (PCC)	Wide application: PCC; Panda, MI,..	Density-based	Density-based	Pearson correlation		FGES + Markov blanket	Bayesian estimation
Weighted	Y	Y	Binary	Binary	Y		Y	Y

FUTURE DIRECTIONS

Perspective

- To improve precision medicine, we need to better understand the complex relationships that exist between different nodes (i.e. genes) and nodes' products in individual samples.
- Networks are a natural way to represent these complex interactions
- Methods to infer networks generally "average" over the members of a population.
- Hence, using networks in precision medicine requires methods that allow inference of network models specific to each individual

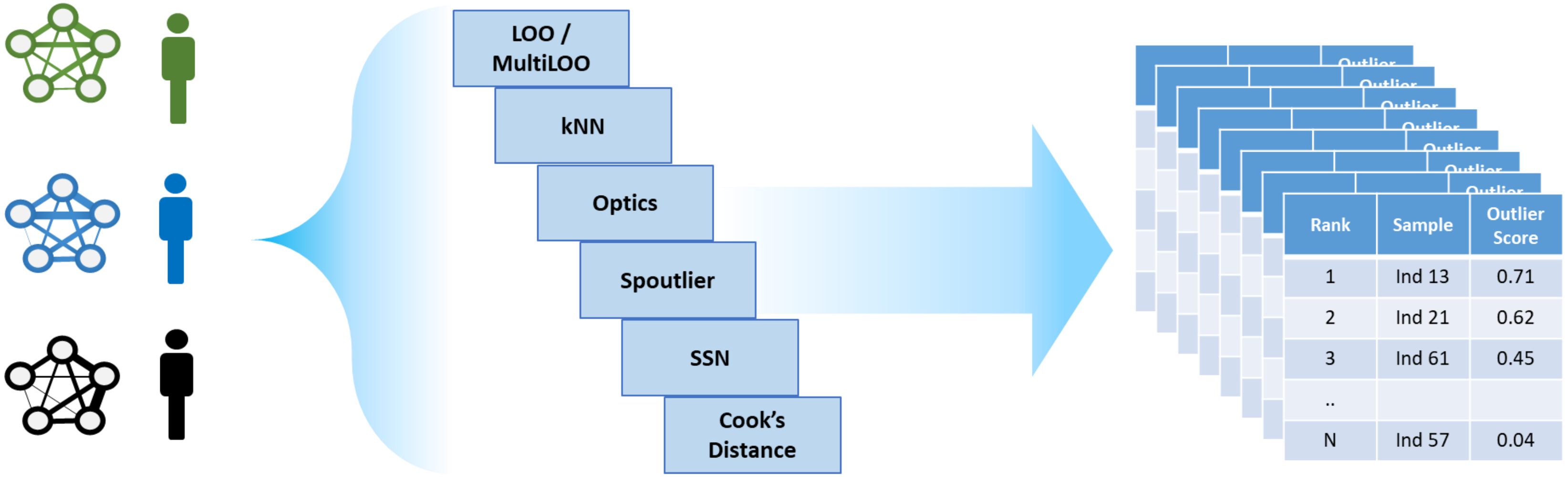
FUTURE DIRECTIONS

Perspective

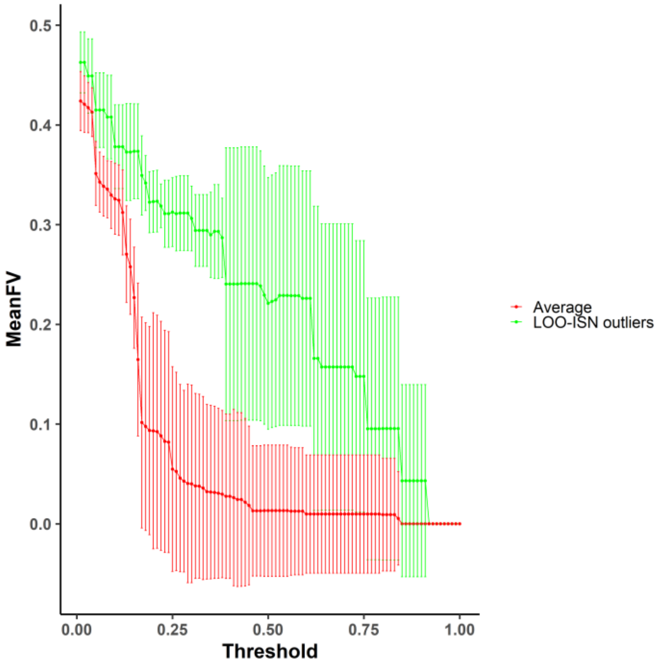
- ISN has been applied consistently results in many fields (transcriptomics, single-cells, microbiome,..) and for many diseases (cancer, covid-19, relapse)
- Applicable both for clustering and prediction tasks
- However, the reported significance assessment has been criticized (Jahagirdar S et al. 2021) for their poor power
- We need for a modular vision to do significance assessment.
  - I.e., consider a module, a set of strongly interacting nodes, to do significance assessment

Modular significance assessment

Future directions



Rank	Sample	Outlier Score
1	Ind 13	0.71
2	Ind 21	0.62
3	Ind 61	0.45
..		
N	Ind 57	0.04

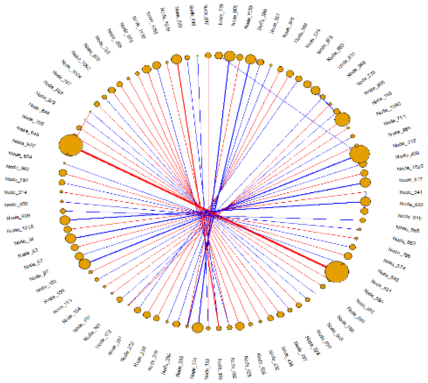


Conclusions

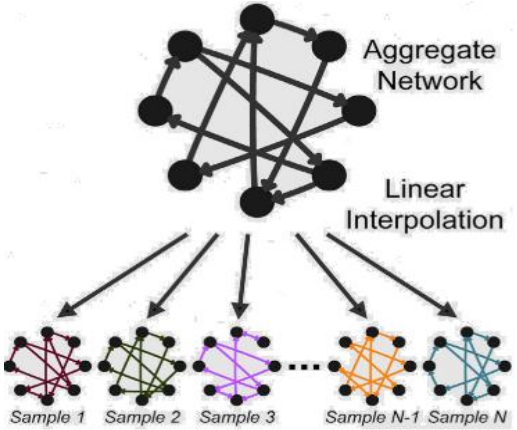
- ISN is an exciting field that promise to complement current information in network analysis
- It is widely applicable in many situations
- Many more challenges to tackle!

SUPPLEMENTARY

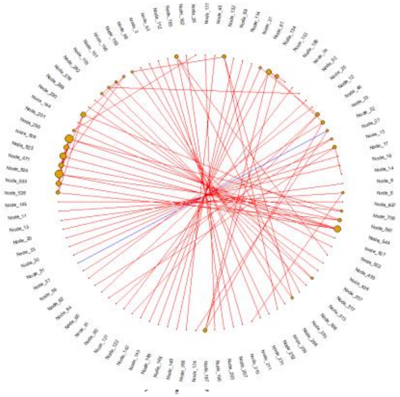
PROJECT SUMMARY



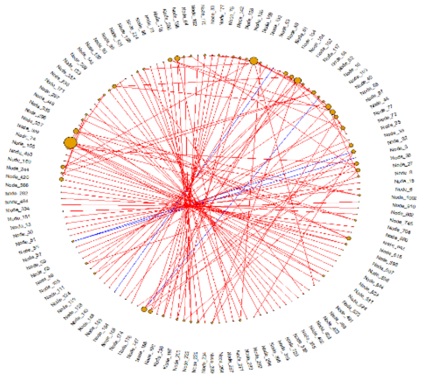
Population-level interaction networks



Adapted from Kuijjer et al. (´19)



Individual 1



Individual 2

Individual-level interactions